Development and Utilization of Genetic Diversity Based Ethiopian Chickpea (*Cicer arietinum* L.) Germplasm Core Collection for Association Mapping

THESIS SUBMITTED TO

Haramaya University, Dire Dawa Ethiopia

For the Award of Doctor of Philosophy

In Plant Breeding

KEBEDE TESHOME KIBRET

International Crops Research Institute for the Semi-Arid Tropics

Patancheru, 502 324

Andhra Pradesh, India

icrisat@cgiar.org

January 02, 2011

Rajeev K. Varshney, *Ph.D. FNAAS*

Director *– Centre of Excellence in Genomics,*
Theme Leader *- Comparative and Applied Genomics (CAG)*
Generation Challenge Program (www.generationcp.org)

Applied Genomics Laboratory

RP-Grain Legumes
Tel: 040 –30713305, 9949994070
e-mail: r.k.varshney@cgiar.org

To Whomsoever It May Concern

This is to certify that research report on "**Development and Utilization of Genetic Diversity Based Ethiopian Chickpea (*Cicer arietinum* L.) Germplasm Core Collection for Association Mapping**" is a bonafide record of work done by Mr. Kebede Teshome Kibret under my supervision and submitted to Haramaya University, Dire Dawa Ethiopia.

Dr. Rajeev K. Varshney

Director – Center of Excellence in Genomics

Theme leader- SP2- Genomics towards Gene Discovery (GCP)

Generation Challenge Program

ICRISAT, Patancheru 502 324

Andhra Pradesh, India.

Tel: +91 (40) 30713071 Extension 2305

Fax: +91-(40)-30713074/ 30713075

Title: Development and Utilization of Genetic Diversity Based Ethiopian Chickpea (*Cicer arietinum* L.) Germplasm Core Collection for Association Mapping

Name: Kebede Teshome Kibret

Institute: Haramaya University, Dire Dawa Ethiopia

Supervisor: Dr. Rajeev K. Varshney

Submitted: January 02, 2012

ABSTRACT

Chickpea (*Cicer arietinum L)* is one of the most important cool season grain legume crops grown in semi-arid tropics and Mediterranean regions. Terminal drought stress is one of the limiting factors for chickpea production. Utilizing of germplasm collections are the main gateway to improve the stagnant production of chickpea in semi arid tropics.

Hence, the objectives of this study were to i) Preliminary phenotyping and genotyping of germplasms collections for diversity assessment; ii) Development of chickpea core collection based on diversity analysis; iii) Identification of desirable accessions for drought tolerance from core set by proper phenotyping; iv) Large scale genotyping of the core collections by SNP markers; v) Large scale genotyping of the core collections by SNP markers; vi) Identification and establishing marker trait associations using appropriate association genetic approaches; vii) Quantification of population structure and relationship of Ethiopian chickpea collection.

The phenotypic evaluation in contrasting environment and SNP marker data analysis revealed that there is significant phenotypic and genotypic variability in Ethiopian chickpea germplasm for drought tolerance and other agronomic traits. The population structure and relationship analysis also revealed strong subpopulation fixation and differentiation which was significantly different from the original population. High allelic and gene diversity were observed in the entire collection with common and rare alleles. Trait marker association analysis showed markers which are strongly associated with maturity related traits and high linkage disequilibrium observed for the polymorphic markers.

Core collection for Ethiopian chickpea germplasm were developed and validated for different validation parameters such as percent mean difference (MD %), percent variance difference (VD %), analysis of variance, coincidence rate of range (CR %), variable rate of coefficient of variance (VR %) and genetic diversity index. The result of validation showed better correspondence between the core set and the entire set which had avoided germplasm duplication and representing the whole collection economically in time and money with few numbers of accessions. Drought tolerant accessions were also identified in the preliminary field screening which needs further confirmation.

DECLARATION

I do hereby declare that the research work presented in this thesis entitled upon "**Development and Utilization of Genetic Diversity Based Ethiopian Chickpea (*Cicer arietinum* L.) Germplasm Core Collection for Association Mapping**" is an original and independent record of research work undertaken by me under the supervision of Dr. Rajeev K. Varshney, Director, CEG (Center of Excellence in Genomics), at International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India, during the period of my study as a part of Ph.D. dissertation research work in plant breeding.

Hyderabad Kebede Teshome Kibret

Date:

Table of Contents **Pages**

List of Tables

List of Figures

Acknowledgements

It is sometimes difficult to express gratitude in few words because the feeling go beyond these. These three years of my Ph.D. work in Ethiopia and ICRISAT- India have changed the course of my thoughts and I have rediscovered myself with the help of some brilliant people I have been associated with.

Firstly, I would like to thank my brilliant supervisor, director Dr. Rajeev K. Varshney for his excellent and unlimited guidance, encouragement, motivation, monitoring and financial support throughout the course of this study.

I wish to thank my other supervisors, Principal scientist Dr Pooran Gaur, Asso. Prof. Firew Mekibib, and Principal scientist Dr Seid Ahmed Kemal for their guidance, encouragement and support throughout the study period.

I would like to extend my thanks to senior scientist Dr Abhishek Rathory, Dr Mahender Thudi and Mr Anil Kumar helping in data analysis and continues encouragement.

I am very grateful to Dr Rozan Mulla and other learning system unit members helping me in facilitating my arrival, encouraging every time in ICRISAT and my safe departure to my country. I appreciate the friendship, guidance of all my lab colleagues, in particular to Mr Abdul Gafoor, Mr Aziz and Mr B.J. Mose for their help in laboratory and green house work.

I am also very grateful to ICRISAT, Sirinka Agriculture Research Center, Wollo University and Haramaya University supporting this finding financially and allowing me to work and stay throughout the course of the study in Ethiopia. I thank Institute of Biodiversity and Conservation (IBC), Ethiopia for giving enough germplasm for this research work especially Dr Tasfaye Wasso who helped me in getting planting material within short period of time.

I am indebted to my family, in particular my wife Zenebech Mekonen and my sister Sehin Desalegn supporting me in every ups and downs, encouraging me in every time to finish this work on time.

List of Abbreviations

ECuA = Ethiopian Custom Authority

LD = Linkage Disequilibrium

SNP = Single nucleotide polymorphism

VD = Variance Difference

MD = Mean Difference

CR = Coincidence Rate

VR = Variable rate of coefficient of variance

DM= Days to maturity

MB = Biomass per plot

DF = Days to 50 % flowering

YLD = Yield per plot

HSW = Hundred seeds weight

HI = Harvest index

PFP = Pod filling period

PPP = Pod per plant

YPP = Yield per plant

PHT = Plant height

SPP = Seeds number per plant

SNNP = Southern Nation and Nationality people

KASPar = KBioscience Competitive Allele Specific Polymerase Chain Reaction

HWE = Hardy Weinberg Equilibrium

GLM = General Linear Model

MLM = Mixed Linear Model

TDS = Tolerant to Drought Stress

DTE = Drought Tolerance Efficiency

DSI = Drought Susceptibility Index

MP = Mean Productivity and RP = Rate of Productivity

1. Introduction

Chickpea (*Cicer arietinum* L.) is the largest produced food legume in South Asia and the third largest produced food legume globally, after common bean (*Phaseolus vulgaris* L.) and field pea (*Pisum sativum* L.) (Gaur *et al.*2010). It is cultivated on 11.55 million hectares of land with annual production of 10.46 million tones with productivity of 955 kg/ha (FAO, 2009).

Ethiopia is the largest producer of chickpea in Africa accounting for about 63 % of the continent's production. It is the second most important pulse crop after faba bean in production and third in area coverage after faba bean and field pea that contributes 16 % of the total pulse production in the country. The total annual chickpea production is estimated about 312080 tones and the national average chickpea yield is 1.33 t/ha (FAO, 2009). The crop also fetches high domestic and export market potential and earning about 22 million US dollar annually (ECuA, 2008).

Chickpea is valued for its nutritive seeds with high protein content. Seeds are eaten fresh as green vegetables, parched, fried, roasted and boiled, as snack food, sweet and condiments and flour as soup. It is grown in Ethiopia with 95 % *desi* and 5% *kabuli* type with different values.

Despite the growing demands and high yield potential (up to 4 t/ha under farmers conditions using new technologies, early planting and disease resistant chickpea varieties, chickpea yields are stable and productivity is stagnant and low (1.3 t/ha). Due to insufficient, untimely and erratic rainfall in these arid and semi-arid areas, the crop often suffers from terminal stress at the end of the cropping season. Terminal drought is globally the number one constraint for production of chickpea and other crops as well. Apart from abiotic stresses, biotic factors like fungal diseases (wilt, root rots and *Ascochyta blights*), African pod borer and storage pests affect chickpea production in most chickpea growing areas.

Until recently, only few germplasm accessions are used in most of the breeding programmes which has led to lower genetic base of cultivated germplasm. This is because most of the available germplasm lines are not yet characterized both at phenotypic and molecular level. The importance of increased use of genetic resources to enhance the genetic potential of the crop for yield and in alleviating the biotic and abiotic stresses has been well recognized (Singh, 1987).

The development of core and mini-core collections has been suggested as a gateway to the utilization of genetic diversity in crop improvement (Upadhyaya and Ortiz, 2001). The core subset would be designed to minimize repetitiveness within the collection and it should represent the rich genetic diversity of a crop. The core collection could serve as a working collection which could be extensively examined, and the accessions which are not included in the core subset would be designated a reserve collection (Frankel, 1984). The information derived from extensive studies on the core subset could be used to guide more efficient utilization of the much larger reserve collection (Tohm *et al*. 1995; Brown, 1989b).

The management and evaluation of large germplasm collections is expensive and inefficient due to redundancies and/or duplications and the impossibility of analyzing with detail all the accessions conserved, particularly in crop plants. Thus, collection management can be significantly improved if the characterization and evaluation steps are focused on a subset of individuals, denominated 'core collection', that represent the diversity conserved in the whole germplasm collection. Molecular markers are proven to be indispensible for the development of core collections in various crop plants.

More recently, association mapping has been applied for different crops like barley where breeding activity has resulted in a high degree of population substructure. It is a method for detection of gene effects based on linkage disequilibrium (LD) that complements QTL analysis in the

development of tools for molecular plant breeding and it addresses false association between markers and phenotypes (Nordborg and Tavare, 2002; Risch and Merikangas, 1996).

We proposed to use the SNP molecular markers, phenotypic and passport data for characterizing and developing a core collection of Ethiopian chickpea germplasm collections.

Hence, the objectives of this study were:

1. Preliminary phenotyping and genotyping of germplasms collections for diversity assessment
2. Development of chickpea core collection based on diversity analysis.
3. Identification of desirable accessions for drought tolerance from core set by proper phenotyping
4. Large scale genotyping of the core collections by SNP markers
5. Quantification of Linkage Disequilibrium using molecular marker data
6. Identification and establishing marker trait associations using appropriate association genetic approaches.
7. Quantification of population structure and relationship of Ethiopian chickpea collection

2. LITERATURE REVIEW

2.1. Origin and Cytology of Chickpea

Chickpea (*Cicer arietinum L.*) is an old world grain legume associated with the Neolithic origin of Near Eastern agriculture (Lev-Yadun *et al.* 2000). Chickpea, also known as *Bengal gram* (Indian), Chickpea (English), *Garbanzo* (Latin America), *Hommes, Hamaz* (Arab world), *Nohud*, *Lablabi* (Turkey), *Shimbra* (Ethiopia), is an edible legume crop. It is the only widely cultivated species of the genus *Cicer* and belongs to the subfamily Faboidae of the Fabacea family (Kupicha, 1981).

In a report by Vavilov (1926), Southwest Asia and the Mediterranean were identified as the two primary centers of origin of chickpea, while Ethiopia as a secondary center of origin. Later, Singh (1997) reported that chickpea most probably originated in region of present day Southeastern Turkey and adjoining areas of Syria. Regarding the origin of kabuli and desi types of chickpea, it is reported that *desi* originated first followed by *kabuli* type which was developed by selection and mutation (Singh, 1987). There is linguistic evidence that *kabuli* type reached India *via* the Afghan capital *Kabul* about two centuries ago and acquired the name as *Kabuli* (van der Maesan, 1987).

The two main types of chickpea are grown widely in the world: '*Kabuli*' (large ram-shaped, cream colored) and '*Desi*' (small angular and dark colored) (Van der Maesen, 1972). The '*Kabuli*' types are grown in the Mediterranean region and the 'Desi' types mainly in the Indian subcontinent. Chickpea is the only widely cultivated species of the genus Cicer.

Availability of living material is a major constraint in some of the wild species of *Cicer* to carry out cytological studies and hence most of the study has been performed on a limited number of *Cicer* species. The chickpea posses a chromosome number 2n = 16 which is highly conserved in different cultivated types, and it has also been characterized with respect to nuclear DNA content. The plant is known to have nine annual and 34 perennial species (van der Maesen, 1987). Study on seven annual species of *Cicer* revealed that these species differ from each other in definite karyotypic features.

2.2. Chickpea Production and Climatic Requirement

2.2.1. Production and Importance

Chickpea is the largest produced food legume in South Asia and the third largest produced food legume globally, after common bean (*Phaseolus vulgaris* L.) and field pea (*Pisum sativum* L.) (Gaur *et al.* 2010). It is cultivated on 11.55 million hectare and annual production of 10.46 million tones with productivity of 955 kg/ha (FAO, 2009). The majority of this area is concentrated in Asia (10.4 million ha) with a production of 9.3 million tones followed by Africa (0.548 million ha) with a production of 0.494 million tones and Americas (0.3 million ha) with a production of 0.4 million tons (FAO, 2009). The ten most important chickpea producing countries are India, Turkey, Pakistan, Iran, Mexico, Mynamar, Ethiopia, Australia, Spain and Canada with 90 % cultivation and consumption from developing countries; of which India accounts for over 67% of the total global chickpea production (FAO, 2009).

Global chickpea production has more or less remained constant since the 1960s. There has been a decline in the area sown to chickpea in India and Pakistan, but this decline was compensated for by

a rise in production in Turkey and, more importantly new producers such as Australia and Canada (Kumar and Abbo, 2001).

Ethiopia is the largest producer of chickpea in Africa accounting for about 63.16 % of the continent's production. It is also the seventh largest producer worldwide and contributes about 3 % to the total world chickpea production (FAO, 2009). It is the second most important pulse crop after faba bean in production and third in area coverage after faba bean and field pea that contributes 16% of the total pulse production. The total annual chickpea production is estimated about 312 thousand tones and the national average chickpea yield in Ethiopia is low, usually 1.3 t/ha (FAO, 2009). The crop also fetches high domestic and export market potential and earning about 22 million US dollar annually (ECuA, 2008).

Chickpea is valued for its nutritive seeds with high protein content. Seeds are eaten fresh as green vegetables, parched, fried, roasted and boiled, as snack food, sweet and condiments and flour as soup. Raw chickpea seeds contain per 100g: 357 calories, 4.5-15.69% moisture, 14.9-24.6g protein, 0.8-6.4% fat, 2.1-11.7g fiber, 2-4.8g ash, 140-440mg calcium, 190-382mg phosphorous, 5-23.9 mg iron, 0-225mg beta-carotene equivalent, 0.21-1.1mg thiamin, 0.12-0.33mg riboflavin and 1.3-2.9 mg niacin (Duke, 1981). It is a major export commodity in Australia ($66 million) and North America ($45 million) during 2005 (FAO, 2006).

Despite the growing demands and high yield potential, chickpea yields are stable and productivity is stagnant at unacceptably low levels. Due to insufficient, untimely and erratic rainfall in these arid and semi-arid areas, the crop often suffers from terminal drought which delays flowering and affect yield at the end of the cropping season. Terminal drought is globally the number one constraint for production of chickpea and other crops as well. Drought causes a considerable (50 %) annual yield loss of chickpea (Varshney *et al*. 2009).

In the past, breeding efforts to improve terminal drought tolerance have been hindered due to its quantitative genetic basis and poor understanding of the physiological basis of yield in water-limited conditions.

Apart from abiotic stresses, biotic factors like fungal diseases and aphids affect chickpea production. The main fungi that affect chickpea are *Fusarium oxysporum sp. ciceri*s causing the plant to wilt and *Ascochyta blight* caused by *Ascochyta rabie*. Ascochyta blight is the most series disease in North India, Pakistan, U.S.A., Africa and the Middle East sometimes causing 100 % yield losses (Pande *et al*. 2005).

2.2.2. Climatic Requirement

Chickpea is usually grown as a rainfed cool-weather crop or as a dry climate crop in semi-arid regions (Muehlbauer and Tullu, 1997). Two main types 'market classes' are recognized within the cultivated chickpea. The *kabuli* type is generally grown in temperate regions whereas the *desi* type is grown in the sub-arid tropics (Muehlbauer and Singh, 1987). The optimum conditions for growth have been suggested to be 21.1-26.7^{0}C day and 17.8-21.17^{0}C night temperatures and an annual rainfall of 600-1000 mm (Kay, 1979, Gaur *et al*. 2010). Chickpea is sensitive to high temperature (> 35^{0}c day light) and to low temperature (< 15^{0}C) which leads to flower drop and reduced pod setting at the stage of reproduction (Gaur *et al*. 2010).

2.3. Drought Stress and Resistance Mechanisms

2.3.1. Drought Stress

Drought is one of the most economically important abiotic constraints to crop production in the world (Araus *et al*. 2002). Low water availability is one of the major causes for crop yield reductions affecting the majority of the farmed regions around the world. Drought can be defined as below normal precipitation that limits plant productivity in the growing season. Five distinct categories of drought affecting crop production in the dry lands, depending on the time of occurrence of drought and general climatic conditions of the region (Hafid *et al*. 1998).

A drought situation can be classified as early season, mid season, late season or terminal, apparent and permanent drought (Kramer and Boyer, 1995).

The early season droughts occur in association with the delay in commencement of sowing rains. Characterization of early season droughts in any agro-climatic region requires precise information on optimum sowing periods for the different crops and their varieties grown in the region under rainfed conditions, amount of rainfall needed to complete the sowing in a given region and the initial amount of rainfall required for safe germination and establishment of the crop stand to minimize the adverse effect of dry spells immediately after sowing.

Mid-season droughts occur in association with the breaks in the monsoon season. If the drought conditions occur during the vegetative phase of crop growth, it might result in stunted growth, low leaf area development, and even reduced plant population. Mid season droughts for crops grown under rainfed conditions can be characterized by the relationship between leaf area index and water use of the crop, depending on the water availability to the crop, and the relationship between the actual leaf area index and effective leaf area index of the crop under moisture stress conditions.

If the crop encounters moisture stress during the reproductive stage because of early cessation of the rainy season, there may be an increase in temperature, hastening the process of crop development to forced maturity. Therefore, late-season droughts have to be characterized on the basis of the relationship between water availability to the crop during the reproductive stage of crop growth and grain yield.

Rainfall in the region may be adequate for one crop but not for others. Therefore, apparent drought conditions are encountered because of mismatching of the cropping patterns to the rainfall/moisture availability patterns in some of the regions.

Drought is a recurring feature in arid regions, as it is in virtually all climate regimes. Even the drought-resistant crops grown in these regions are likely to be subjected to moisture stress, even during years with above-normal rainfall. Alternate land use systems have to be introduced in these regions for sustainable agriculture.

2.3.2 Drought Stress Resistance Mechanisms

Water deficits result from low rain fall, poor soil water storage and when the rate of transpiration exceeds water uptake by plants. Yield reduction due to drought ranged from 30 to 60 percent in chickpea which depends on geographical region and length of crop season. Since drought is accompanied by relatively high temperature, which promotes high evapotranspiration and hence could accentuate the effects of drought, the yield reduction will increase more than this in some parts of chickpea producing areas (Sabaghpour *et al*. 2006).

Although plant tolerance is an important objective in many plant breeding programs, understanding of the physiological mechanisms that contribute to variability in crop performance in drought environments remains limited (Passioura, 1996). Many physiological processes associated with crop growth and developments including CO_2 assimilation, transpiration and

stomatal regulation, cell growth, hormonal and enzyme concentration *etc*. are influenced by water deficits (Turner and Begg, 1978). The essence of good drought management is to use the crop ranges of response to best advantage.

Plants have developed various strategies to acquire stress tolerance. These strategies include changes in metabolic processes, structural changes of membranes, expression of specific genes and production of secondary metabolites. In genetic sense, the mechanisms of drought resistance can be grouped into three categories, drought escape, drought avoidance and drought tolerance. However, crop plants use more than one mechanism at a time to resist drought (Mitra, 2001).

Drought escape is the ability of a plant to complete its lifecycle before serious soil and plant water deficits develop. The plants can escape from drought by early flowering and maturity before the stress occur (Turner, 1979). Xu *et al*. (2005) studied quantitative trait loci (QTL) for drought escape and tolerance in set of introgression lines of rice, they found twelve main-effects QTL (M-QTLs) for heading days and mapped to ten rice chromosomes except chromosomes 2 and 11. In addition, five pairs of epistatic QTL (e-QTLs) affecting heading days were identified including two pairs detected under the irrigated condition, one pair under stress and two pairs by the heading days differences across water levels.

Drought escape can be defined as the ability of a plant to complete its life cycle before a serious plant water deficit develops. Selection for rapid phonological development is a common approach in breeding for drought resistance in crop (Jordan *et al*. 1983).

Quisonberry (1982) defined resistance as the ability of a plant variety to produce a higher yield than another at a given limiting level of water availability. Crop adaptation mechanism in response to decrease water availability further divided in to drought escape, dehydration tolerance and dehydration avoidance (Verslues *et al*. 2006).

Dehydration avoidance defined as the plant's ability to retain a relatively higher level of water potential under soil and atmospheric water stress (Levitt, 1972). The first response of a plant to stress is limiting water loss mainly by stomatal closure which is linked with reduction in carbon gain by the plant. The other mechanism for the control of water loss include the reduction in radiation load *via* change in plant canopy architecture and change in root and shoot growth as the long term morphological adjustments. Dehydration avoidance is the mechanism of drought tolerance where by plants keep high water potential in the tissue by maintaining water uptake through deep rooted structure and increasing hydraulic conductance or reduction of water loss by means of structural adjustments like stomatal control and reduction in evaporative surface, increasing water use efficiency and absorbed radiation (Mooney *et al*. 1977).

Dehydration tolerance is the survival mechanism when water stress is more severe. The ability of tissue to maintain turgor pressure during severe water stress is an important mechanism of dehydration tolerance (Hsiao *et al*. 1976). It is a type of drought tolerance whereby plants survive at low water potential by solute accumulation and increase elasticity to avoid desiccation. When the plant is exposed to low water potential, it will prepare protective proteins like heat shock proteins, late embryogenesis abundant (LEA) proteins and accumulation of abscisic acid (ABA) (Creelman and Zeevaart, 1985). For agricultural context, drought resistance mechanism related to productivity (drought escape and dehydration avoidance) is very important.

In chickpea, the focus of drought resistance research is on the ability to sustain greater biomass production and crop yield under seasonally increasing water deficit, rather than the physiological aptitude for plant survival under extreme drought shock (Serraj and Sinclair, 2002). This has led to the focus on escape and avoidance strategies such as early maturity (Kumar and Abbo, 2001) and large root systems (Kashiwagi *et al*. 2006).

2.4 Genetic Diversity

The definition of genetic diversity is referring to the variance at individual gene loci (among alleles of a gene), among several loci or gene combinations, between individual plants within plant populations, or between plant populations (Smale and McBride, 1996).

Genetic diversity refers to the variation among alleles of genes in different individuals of population of species. While the ultimate source of genetic diversity is gene mutation. It is molded and shaped by selection, recombination, genetic drift and migration in the face of heterogeneous environment in space and time (Falconer and Mackey, 1996).

Genetic diversity gives species the ability to adapt to changing environments, including new pests and diseases and new climatic conditions, such as global warming (Parmesan and Yohe, 2003). Differences between genotype with regard to agronomic characters, morphological characters, biochemical characters and molecular characters are either indirect or direct representations of differences at DNA level and are therefore expected to provide information about genetic relationships and allelic richness and evenness of the genotypes (Shannon's index). Genetic diversity commonly is measured by genetic distance (GD) or genetic similarity (GS = GD-1), both of each imply that there are either differences or similarities at the genetic level (Weir, 1990). The assessment of genetic diversity is important not only for crop improvement but also for efficient management and conservation of germplasm resources, identification of duplicate accessions in the germplasm and in applied breeding program.

Analysis of genetic relationships in crop species is also an important component of crop improvement programs, as it serves to provide information about genetic diversity, and is a platform for stratified sampling of breeding populations. A detailed knowledge of genetic relationship among accessions is an important factor for the success of plant breeding programs

and for efficient sampling and more enhanced utilization of available germplasm. Assessment of the extent of genetic variability with in chickpea is fundamental for chickpea breeding and conservation of genetic resources and is particularly useful as a general guide in the choice of parents for developing hybrids.

Criteria for the estimation of the genetic diversity can be different, which include morphological traits and molecular markers (Upadhyaya, *et al*. 2007; Rao *et al*. 2007). Molecular markers have proved to be valuable tools in the characterization and evaluation of genetic diversity within and between species and populations. Using these molecular markers, large amount of genetic variation which exists between chickpea genotypes can be used efficiently for gene tagging and genome mapping of crosses to introgress the favorable traits and to genotype large composite collections present in gene bank (Talebi *et al.* 2008).

2.5 Germplasm Management and Utilization

Plant genetic resources (PGR) are the most important components of agro-biodiversity. The plant genetic resource include primitive forms of cultivated plant species and landraces, modern cultivars, obsolete cultivars, breeding lines and genetic stocks, weedy types and related wild species (IPGRI, 1993). The plant genetic resources contribute enormously towards achieving the Millennium Development Goals of food security, poverty alleviation, environmental protection and sustainable development.

Over the years, gene banks have been established in a number of countries and the number of accessions conserved in about 1400 gene banks now exceeds six million (FAO, 1998). This increase in accession numbers in gene banks and lack of corresponding increase in their use by the crop improvement scientists was a clear indication that the collections were not being used to their full potential (Marshal, 1989).

The management of genetic resources is a complex, multi-faceted process. It involves a number of distinct stages, which are nevertheless linked and interrelated, from the selection of priority taxa, to the design and implementation of complementary conservation strategies and the development and exchange of the results of germplasm use (Maxted *et al.* 1997a).

In general, plant genetic resource (or simply germplasm) management comprises two phases. The first, germplasm conservation, includes acquisition, or securing germplasm in situ (by establishing reserves) or ex situ (by assembling collections through exchange or exploration). It also comprises maintenance: monitoring and protecting germplasm in reserves or storing it ex situ under controlled conditions, propagating it while preserving its original genetic profiling with maximum fidelity, monitoring its viability and health in storage and maintaining associated passport and other data. Germplasm conservation also involves characterization, assaying highly heritable morphological and molecular traits of germplasm, for taxonomic, genetic, quality assurance and other management purposes (Janick, 1995).

The second phase of germplasm management encouraging utilization includes evaluation, assaying germplasm for agronomically or horticulturally meritorious traits with relatively low heritability and high components of environmental variance (high yield, adaptation and resistance to stress). Genetic enhancement or making particular gene more accessible and usable to breeders by adapting exotic germplasm to local environment without losing its essential genetic profile or introgressing high value traits from exotic germplasm to adaptive varieties.

2.6 Developing Core Collections

Some years ago, Otto Frankel suggested that forming core collections was a way to meet the challenge of the growing sizes and numbers of collections of plant genetic resources (Frankel,

1984). He did so at a biotechnology symposium where it was clear that the emergence of molecular biology would have a significant impact on germplasm collections.

Genetic resources stored in gene banks are usually sampled to foster efficient evaluation and utilization of the collections as well as to study phenotypic and genotypic diversity, from core subsets, and eliminate redundant and duplicate accessions within accessions. The main purpose of developing core collection is to preserve in the sample as much of the diversity present in the original collection as possible (Crossa *et al*. 1995a). For example, the approach of forming core collections (core subsets) was introduced to increase the efficiency of describing and using collections stored in gene banks, while preserving as much as possible the diversity of the entire collection (Frankel and Brown, 1984; Brown, 1989).

The process of sampling genetic resources with the objective of forming subsets starts with grouping accessions to obtain homogenous within and heterogeneous between clusters (or groups) and then using a predetermined sampling strategy within each cluster. The grouping of accessions in to clusters is achieved by a classification strategy that partitions the original collections in to groups with maximum distances between accessions located in different groups and minimum distances between accessions located in the same group Franco *et al.* (1998, 1999, 2002) and Franco and Crossa (2002) proposed a sequential Ward-Modified Location Method (MLM) the strategy in which the Gower (1971) distance is used as a measure of similarity (or distance among accessions considering all continues and categorical variables.

The sampling intensity of core collection development ranged from 5 % to 20 % of the total number of accessions. This intensity of sampling captures 86-90 % of the diversity present in the reserve collections (Brown, 1989).

On the basis of several statistical model, Brown (1989a, 1989b) suggested that at least 70 % of the alleles present in the entire collection will be represented in a core collection comprised at least 10 % of the accessions. Large increases in core collection size have increasingly marginal effects on the levels of diversity retained (Brown, 1989a). For example, under conditions of variables levels of diversity in a population of 10000, about 70 % of alleles were predicted to be retained in core comprised of 10% of the accessions, but doubling the number of retained accessions to 20 % increase the predicted diversity retention by only about 5 % (Brown, 1989a). This result suggested that a core collection comprised of 10 % of the accessions is nearly as efficient as much larger collection in representing allelic diversity.

Compared with the traditional plant breeder, the molecular plant breeder would need access to a wider diversity of genetic resources, but not in great numbers of accessions. In addition to these emerging currents, a new array of technologies such as tissue culture, cryopreservation and DNA libraries were coming on stream as new options for germplasm storage. Such methods could help to avoid costly recurrent regeneration, but again might not be sensible to implement on large numbers of samples.

One of the reasons that traditional plant breeders are using less basic germplasm in research is the lack of information on traits of importance, which often show high genotype x environment interactions and requires replicated multi-location evaluation. This is a very costly and resource-demanding task owing to the large size of the germplasm collections.

Frankel was concerned that large germplasm collections might be stifled by their own apparent success. Thus at a time when many were clamoring for more collecting, he put forward the radical alternative that fewer, smaller collections were better.

To overcome this, studying, the diversity of germplasm collection and developing 'core collections', which are about 10 % of the entire collection, but attempt to maximize the diversity represented, is proposed by Frankel (1984). He proposed 'core collection which would represent with a minimum of repetitiveness the genetic diversity of a crop species and its relatives.

When the size of the entire collection is very large even a core collection size becomes large for the breeders to evaluate. To overcome this, ICRISAT scientists have developed a seminal two-stage strategy to develop a mini-core collection that represents the diversity of the entire collection, which includes 10% of accessions of core collection and 1% of the entire collection (Upadhyaya and Ortiz, 2001). This mini core collection still represents the diversity of the entire collection. Due to the reduced size, core collection can be evaluated extensively to identify the useful parents for crop improvement.

A core collection consists of a limited set of accessions derived from a germplasm collection, chosen to represent the genetic spectrum in the whole collection, and including as much as possible of its genetic diversity (Brown, 1995). The core collection provides a focus for effort that is for a combined effort of gene bank workers, breeders and other researchers. Its purpose is to attract multiple use and many users.

Four elements are basic to the concept of a core collection

1. The parent whole collection is a large entity (from the stand point of management and use of many accessions)
2. The core from this large collection has a restricted size (5- 20 %)
3. The core is the representative sample of the collection
4. It is diverse

Typically there are four stages in forming a core collection. These are:

1. Defining the collection to be represented, assembling all the relevant data on the accessions in that collection and deciding the size of the core

2. Grouping the accessions into groups that reflect the major genetic and ecological categories within the whole collection

3. Choosing the entries for the core - how many per group and which accessions

4. Managing the core set.

2.7 Genetic Markers

Genetic markers represent genetic differences between individual organisms or species. Generally, they do not represent the target genes themselves but act as 'signs' or 'flags'. Genetic markers that are located in close proximity to genes (i.e. tightly linked) may be referred to as gene 'tags'. Such markers themselves do not affect the phenotype of the trait of interest because they are located only near or 'linked' to genes controlling the trait. All genetic markers occupy specific genomic positions within chromosomes (like genes) called 'loci' (singular 'locus').

There are three major types of genetic markers: (1) morphological (also 'classical' or 'visible') markers which themselves are phenotypic traits or characters; (2) biochemical markers, which include allelic variants of enzymes called isozymes; and (3) DNA (or molecular) markers, which reveal sites of variation in DNA (Winter & Kahl, 1995).

2.7.1 Morphological Markers

Morphological markers are traditional markers widely used by plat breeder before the invention of molecular markers. Morphological markers are usually visually characterized phenotypic characters such as flower color, seed shape, growth habits or pigmentation. Morphological markers are very few simple Mendelian morphological characters which have been discovered in

forest trees that could be used as genetic markers. Many of the identified morphological markers are mutations observed in seedlings such as albino needles, dwarfing and other aberrations. Such mutants have been used to estimate self-pollination rates in conifers. These markers, however, have limited application because morphological mutants occur rarely and often are highly detrimental or even lethal to the tree.

2.7.2. Biochemical Markers

Allozymes have been the most important type of genetic marker in forestry and are used in many species for many different applications. Allozymes are allelic forms of enzymes that can be distinguished by a procedure called electrophoresis. The more general term for allozymes is isozymes, and refers to any variant form of an enzyme, whereas allozyme implies a genetic basis for the variant form. Most allozyme genetic markers have been derived from enzymes of intermediary metabolism, such as enzymes in the glycolytic pathway; however, conceivably an allozyme genetic marker could be developed from any enzyme ((Adams *et al.* 1992).

Isozymes were defined as structurally different molecular forms of an enzyme with, qualitatively, the same catalytic function. Isozymes originate through amino acid alterations, which cause changes in net charge, or the spatial structure (conformation) of the enzyme molecules and also, therefore, their electrophoretic mobility. After specific staining the isozyme profile of individual samples can be observed (Soltis & Soltis, 1989).

2.7.3. Molecular Markers

Molecular markers reflect heritable differences (e.g. polymorphisms) in homologous DNA sequences among individuals. These differences may be due to single nucleotide polymorphisms (SNPs), insertions or deletions or rearrangements (translocations or inversions). The methods of

detection of polymorphism involve the use of restriction endonuclease, nucleic acid hybridization or DNA sequence amplification.

A large number of reviews have been published on molecular markers and their application in crop improvement (Lörz and Wenzel, 2005). The decision which marker system is the most appropriate to use will depend on the species, the objective of the marker work and on the resources available. Here the most widely used molecular marker technologies will be described

2.7.3.1 Restriction Fragment Length Polymorphism (RFLP)

It should be recognized that isoenzyme and other protein-based marker systems are the first represent genetic markers and were in wide use long before DNA markers became popular. The concept of using variations at DNA level as genetic markers started with the restriction fragment length polymorphism (RFLP). When the DNA of different individuals is digested with restriction enzymes, differences in size of the resulting fragments of DNA can be visualized via Southern hybridization with labeled probe. The differences are due to evolutionary changes in sequence of nucleotides in the DNA of different individuals. The first documentation of RFLP came from viruses followed by subsequent elegant demonstration made in the human globulin gene cluster (Jeffreys, 1979). Since then, most organisms have been explored for the presence of RFLP and application of technology has evolved in various fields.

RFLPs are differences in restriction fragment lengths caused by SNPs or INDELs that create or destroy restriction endonuclease recognition sites. Both the basis and techniques for RFLPs (Botstein, 1980) in plant genome mapping have been extensively reviewed (Tanksley *et al*. 1989). RFLPs are assayed by hybridizing labeled (c) DNA probes to a Southern blot (Southern, 1975) of genomic DNA digested with various restriction enzymes. Marker alleles are identified by size differences of the restriction fragments to which these probes hybridize. The RFLP marker

technology allowed the construction of the first whole-genome linkage maps in plants (Bernatzky and Tanksley, 1986; Helentjaris *et al*. 1986) and initiated the rapid developments in the field of comparative genomics (Gale and Devos, 1998; Paterson *et al*. 2000).

Some advantages of the use of RFLPs are that, if single-copy, most markers can be scored co-dominant, are locus specific and high-throughput PCR-based markers can easily be developed from the probe sequences.

Some limitations to the use of RFLPs are:

- Development of RFLP probe sets and markers is labor intensive and the multi-step protocol is time- consuming.
- Analysis requires large amounts (1-10 μg/gel lane) of high-quality DNA.
- RFLPs are difficult to automate/multiplex and therefore have a low genotyping throughput.
- RFLP probes must be physically maintained and thus are difficult to share between laboratories.

2.7.3.2 Random Amplified Polymorphic DNAs (RAPDs)

RAPD markers are defined as DNA polymorphisms produced by "rearrangements or deletions at or between oligonucleotide primer binding sites in the genome" (Welsh and McClelland 1990; Williams *et al*. 1990). The method simultaneously amplifies DNA fragments with a single random-sequence primer (usually 10-base oligomers) in a low-stringency PCR (35-45°C). These fragments are separated on conventional agarose gels and RAPDs are identified by the presence or absence of a particular fragment (i.e. band). RAPD markers can be converted into simple and robust PCR markers termed Sequence Characterized Amplified Regions (SCARs) by developing site specific primer pairs from cloned RAPD fragments. DNA Amplification Fingerprinting (DAF) is a modified approach of the RAPD technique. It employs one or more primers as short as

five nucleotides in length to produce complex banding patterns that are resolved by polyarcylamide gel electrophoresis.

The major advantage of the use of RAPDs is the use of universal primers (Tingey *et al.* 1994). Other advantages are the small amount of DNA required (5-25 ng/individual) and the relative low start-up costs (Waugh and Powell, 1992). The major limitations to the use of RAPDs are:

- The reproducibility of RAPD assays across laboratories is generally low (Perez *et al*. 1998).
- Most RAPD markers are dominant, although some can be converted into locus-specific co-dominant markers (Davis *et al*. 2005).
- The homology of fragments across genotypes cannot be ascertained without sequencing.

2.7.3.3 Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphism (AFLP) is a successful, PCR-based multi-locus fingerprinting technique that efficiently identifies DNA polymorphisms without prior sequence information (Vos et al. 1995; Mueller and Wolfenbarger, 1999). The polymorphisms are scored by differences in restriction fragment lengths caused by SNPs or INDELs in or adjacent to the endonuclease restriction sites. AFLP assays are performed by selectively amplifying a subset of genomic restriction fragments using PCR. The selectivity is achieved by using selective nucleotides that are added to the 3' ends of the PCR primers that anneal to the adapters legated to the restriction sites. Only restriction fragments in which the nucleotides flanking the restriction site match the selective nucleotides will be amplified. The subset of amplified fragments is then separated with gel electrophoresis to generate the fingerprints.

The development of the AFLP method has had a large impact on genomics as it was the first method that cost-effectively enabled the identification and typing of a large number of markers throughout the genome using a simple and robust protocol (Blears *et al*. 1998). A major

improvement has been made by switching from radioactive to fluorescent dye-labeled primers for the detection of fragments in gel-based or capillary DNA sequencers (Schwarz *et al*. 2000). The success of the technology mainly can be contributed to the high multiplex ratio and genotyping throughput, the high reproducibility, the low amount of DNA required (200- 500 ng) and due to the fact that it can be applied to virtually any organism without prior sequence information.

Some limitations to the use of AFLPs are:

- Scoring of markers is based on fragments that are separated on length by electrophoresis on gel-based systems which limits the throughput.
- The homology of an amplified fragment cannot be unequivocally ascertained across genotypes or mapping populations. This makes that AFLP markers are difficult to use as anchor markers.
- To obtain sufficient genome coverage multiple primer combinations for one restriction enzyme combination need to be analyzed "in serial" for each sample.
- The extraction of an AFLP fragment from a polyacrylamide gel and conversion into a simple PCR marker is a labor-intensive and sequential procedure (Brugmans *et al*. 2003; Polanco *et al.* 2005).
- AFLP is a proprietary technology, owned by Keygene N.V. Because the technology is patented, access is restricted for the commercial use of the technology in certain crop species without prior agreement.

2.7.3.4 Microsatellites

Microsatellites also called Simple Sequence Repeats (SSRs), Short Tandem Repeats (STRs) or Sequenced-Tagged Microsatellite Sites (STMS) are tandemly repeated mono-, di-, tri-, tetra-,

penta, and hexa-nucleotide motifs. SSR length polymorphisms are caused by differences in the number of repeat units. SSR loci are amplified by PCR using pairs of oligonucleotide primers specific to unique DNA sequences flanking the SSR sequence. SSRs tend to be highly abundant and polymorphic and randomly dispersed throughout most genomes (Tautz and Renz, 1984; Lagercrantz *et al*. 1993; Goldstein and Pollock, 1997). SSRs can be identified in genomic libraries or within genes by searching Expressed Sequence Tags (EST) databases (Gupta *et al*. 2003; Thiel *et al*. 2003). The increased availability of large EST datasets and full genome sequences enables the rapid identification of these repeats for some species. Most SSR markers are co-dominant and define a specific locus which is a major advantage in population studies (Thiel *et al*. 2003).

SSR markers also have some limitations:

- The marker discovery phase is expensive and involves DNA sequencing.
- A high resolution gel equipment system is required for genotyping and the throughput is limited as a consequence of its reliance on gel or capillary electrophoresis.
- Developing and optimization of a multiplex SSR assay is labor intensive. Although some degree of multiplexing can be achieved, often SSRs are amplified separately and later pooled for analysis.

2.7.3.5 Single Nucleotide Polymorphisms (SNPs)

The new generation of genetic markers is based on SNPs, which are defined as single nucleotide positions in a given DNA stretch at which variations between different individuals within a species occur. SNPs are single base pair position in genomic DNA at which different sequence alternatives (alleles) exist in normal individuals in some populations, wherein the least frequent allele has an abundance of at least 1 % or greater. SNP is a polymorphism occurring between DNA samples with respect to single base. In general, SNPs are the most common form of DNA

sequence polymorphisms present (Collins *et al*. 1997). The binary (bi-allelic) character and stability from generation to generation make SNPs amenable to automated, high throughput genotyping and, therefore attractive tool for quantitative trait loci (QTL) mapping studies and marker assisted selection (MAS) in plant breeding programs.

The SNPs have become marker of choice. Due to their abundance in genome, they are extremely useful for creating high-density genetic map. This density cannot be achieved with other genetic marker classes. Due to this abundance, SNPs have the potential to provide basis of a superior and highly informative genotyping assay. SNPs in coding regions (cSNPs) may have significance functional if the resulting amino acid change causes the altered phenotype. SNP markers associated with phenotypic changes pinpoint functional polymorphism. At a particular site in a DNA molecule theoretically four possible nucleotides are involved but in reality only two of this four possibilities have been observed at the specific sites in a population, those SNPs are largely biallelic in nature. Although the biallelic nature SNPs makes them less informative per locus examined than multiallelic markers such as RFLPs and microsatellites but this difficulty is overcome by their abundance which allows the use of more number of loci (Xiong and Jin, 1999). SNPs are less mutable as compared to other markers, particularly microsatellites. The low rates of recurrent mutation make them evolutionarily stable. They are excellent markers for studying complex genetic traits and for understanding the genomic evolution.

The "golden standard" for SNP detection has been the Sanger dideoxy-sequencing method. Since this method generates more information than necessary, misses SNPs when the DNA template is heterozygous and, thus is time-consuming and very expensive other gel-based assays were soon developed. These methods include Cleaved Amplified Polymorphic Sequence (CAPS), Allele

Specific PCR (AS-PCR) and Single Stranded Conformation Polymorphism (SSCP) (Suzuki *et al.* 1991; Konieczny and Ausubel, 1993; Prosser, 1993).

In all these methods PCR products are separated on agarose or polyarcylamide gels for SNP determination, limiting the throughput that can be obtained. This forced the development of a wide-variety of high-throughput SNP genotyping platforms that make use of the reaction principles of minisequencing, heteroduplex analysis and allele specific hybridization (Henry, 2001). A number of reviews have been published in which these and other SNP detection methods are discussed extensively (Gupta *et al*. 2001; Syvanen, 2005).

Despite the fact that SNPs are the preferred markers in human genotyping, their application to other species, however, is limited by two important factors:

- Although high-throughput SNP assays are developed, most methods still require a marker specific amplification reaction (e.g. MulltiPlex Ligation dependant Probe Amplification), marker specific primers (e.g. Single Base Extension, SNP Wave, Invader and Pyrosequencing), oligonucleotides (e.g. Padlock assays, Single Feature Polymorphism arrays) or probes (e.g. Taqman, Molecular Beacon).
- The initial investment required for marker discovery and assay development remains prohibitive for many species.

A wide range of marker techniques is now available for genotyping plant genomes. Markers are not only employed in basic research but also, with increasing frequency, in practical plant breeding. The choice of marker system depends on the species, the objective of the marker work and on the resources available. For population studies SSRs are highly informative, but for identification of QTLs and marker assisted selection genome-wide PCR-based markers, like SNPs, are usually the markers of choice. Unfortunately, highly informative marker types like

SNPs and SSRs are currently extensively available only for a limited number of well- studied crop plants. Genotyping in plant or fungal genomes for which little or no genetic resources are available still has to be performed using universal marker techniques like RAPD and AFLP. Although successful, these technologies are restrained by their throughput.

2.7.3.6 Diversity Arrays Technology (DArT) Marker

To overcome some of the above mentioned limitations of currently available marker technologies the Diversity Arrays Technology (DArT) has been developed (Jaccoud *et al*. 2001). DArT is a generic, hybridization-based and cost-effective fingerprinting method. A single DArT assay simultaneously types hundreds to thousands of SNPs and insertion/deletion polymorphisms (InDels) across the genome. The DArT procedure essentially encompasses five steps: 1) construction of a genomic library (i.e. genomic representation), 2) printing of the genomic library on microarrays, 3) labeling of genomic representations, 4) hybridization of this labeled genomic representation on the microarray, followed by washing and 5) scanning and data analysis.

Diversity array technology has several advantages compared to existing molecular marker technologies:

- DArT is capable of parallel instead of serial analysis of marker data. Many marker technologies are constrained by their dependence on gel electrophoresis, resulting in low throughput. On polyacrylamide gels, for example, between 50-150 DNA fragments, can be electrophoretically separated, while array-based methods (e.g. DArT) can accommodate much higher densities and are therefore capable of parallel rather than serial analysis of marker data.
- DArT does not require DNA sequence information. Some of the existing marker methods (e.g. SSRs and many SNP based methods) require sequence information before assays can be

developed. Although new sequencing technologies rapidly emerge, DArT is independent of investments in genome sequencing. DArT therefore is of particular interest for species for which limited or no genetic resources are available (i.e. orphan crops). In addition it remains to be seen if large complex (e.g. polyploid) genomes are amenable to be sequenced.

- DArT markers are scored with high accuracy. Specifically developed software (DArTsoft) analyzes the large amount of data generated in each DArT experiment. The software analyzes the microarray images and subsequently identifies and scores markers as described by Wenzl *et al*. (2004). The program is unique in the fact that it calculates a range of quality parameters (Akbari *et al*. 2006) for each marker. The thresholds for these quality parameters can be set by the user to objectively select a set of markers with high quality and reproducibility.
- DArT is an open source platform. DArT Pty/Ltd. has established a network of DArT users (www.diversityarrays.com/dartnetwork.html), who will contribute their scientific expertise and resources to develop and improve the technology further.
- DArT platform allows flexibility of applications. DArT libraries are prepared from individual or pooled genomes (i.e. meta-genome) of the individuals that best suit the desired application. For mapping studies this often are the parents of the segregating population, but for genetic diversity studies the DNA can be derived from cultivated varieties to wild relatives. The microarray platform itself is flexible as well. In initial experiments markers can be identified in the genomic library (discovery arrays). These markers can then be re-arrayed on new slides (genotyping arrays) and serve for the high-throughput detection of hundreds to thousands of markers in large populations.

2.8 Applications of Genetic Markers

The main advantages of using molecular markers are that they measure the genetic diversity at DNA level, can account for the effects of selection, are environment-independent, and are available in an enormous number. Molecular markers are also used to identify and discriminate between closely related cultivars, to identify the phylogenetic-relationships of crop species, linkage map construction, genome organization and tagging loci affecting quantitative traits (Azhaguvel *et al*. 2006).

Another possible application of molecular markers is in germplasm collections and characterization through genotyping. These applications include identification and verification of old and new collected genotypes; detection of duplicates; genetic purity analysis; genetic diversity analysis; construction of core collection and selection of interesting, gene resources; monitoring of viability and health and genetic changes due to long-term storage at low temperature.

2.9 Association Mapping

One hallmark of twentieth-century genetics will be the tremendous strides made in understanding how individual genes control simple traits (phenotypes). However, the fruits of the revolution in molecular genetics will likely be seen in this century, when the genes and alleles that control complex traits [quantitative trait loci (QTL)] are identified and understood.

For the past decade of this century, there has been success in using conventional map-based strategies in identification and cloning of quantitative trait loci (QTL) in model plant species including tomato and Arabidopsis. These quantitative traits are generally the products of many loci with varying degrees of effect upon the observed phenotypes.

Mapping of genes controlling quantitative traits in plant normally involves the use of segregating populations derived from parents with contrasting phenotypes and/or genotypes. Recombination

frequencies between markers and the genes of interest are estimated from their patterns of co-segregation. But this has the following limitations. Firstly, there is a need to grow two to three generations before linkage analysis to the gene of interest or quantitative trait is possible. Secondly, very large segregating populations are required to achieve high resolution mapping, which may be needed for marker assisted selection (MAS) or cloning of candidate genes by chromosome landing strategies (Tanksely *et al*. 1995), and thirdly, only two alleles at any particular locus can be assessed.

Recently a new approach to genetic mapping has emerged called association mapping. Association mapping is a useful tool for crop genetic improvement that identifies polymorphic markers associated with phenotypic variation for important traits (Flint-Garcia *et al*. 2003). It is a method for detection of gene effects based on linkage disequilibrium (LD) that complements QTL analysis in the development of tools for molecular plant breeding and it addresses false association between markers and phenotypes. It is emerged as a tool to resolve complex trait variation down to the sequence level by exploiting historical and evolutionary recombination events at the population level (Nordborg and Tavare, 2002; Risch and Merikangas, 1996).

As a new alternative of traditional linkage analysis, association mapping offers three advantages, (i) increased mapping resolution, (ii) reduced research time, and (iii) greater allele number (Yu and Buckler, 2006). Since its introduction, it has continued to gain favorability in genetic research because of advances in high throughput genomic technologies, interests in identifying novel and superior alleles, and improvements in statistical methods (Thornberry *et al*. 2001).

Such association studies involving the use of germplasm collections or natural plant populations for the identification of molecular markers linked to QTLs. This whole genome association mapping using diverse germplasm; enables to detect candidate genes, detecting pleiotropic genes

and genes showing interactions with environment. But it requires comprehensive phenotypic data for modeling genotypic x environment interactions.

Association mapping theoretically allows mapping with higher resolution than achieved using bi-parental crosses (Tommasini *et al*. 2007). The degrees of resolution depend on the extent of linkage disequilibrium (Ramington *et al*. 2001) and higher resolution when linkage disequilibrium (LD) declines rapidly with increasing genetic distance. While using the association analysis, the statistical power of associations is determined by the extent of LD with the causative polymorphism, as well as sample size used for the study (Wanga and Rannala, 2005).

2.10 Linkage Disequilibrium

Genetic linkage generally refers to coinheritance of different loci within a genetic distance on the chromosome. There are two terms used in population genetics, linkage equilibrium (LE), and linkage disequilibrium (LD) to describe linkage relationships (co-occurrence) of alleles at different loci in a population. LE is a random association of alleles at different loci and equals the product of allele frequencies within haplotypes, meaning that at random combination of alleles at each locus its haplotypes (combination of alleles) frequency has equal value in a population.

In contrast, LD is a nonrandom association of alleles at different loci, describing the condition with unequal (increase or reduced) frequency of the haplotypes in a population at random combination of alleles at different loci. The principles leading to LD apply to both biparental mapping populations (F2, RILs, etc) and natural populations.

On the one hand, the high level of LD in self pollinated crops is due to the inbreeding mating type of this species; on the other hand, the selection of germplasm plays an important role in analysis

of a germplasm collections, cultivars, land races and wild accession provided hints that the level of LD increases from cultivars to landraces to wild relatives (Caldwell *et al*. 2006).

2.11 Population Structure

Population stratification exists when the total population has been formed by admixture between subpopulations and when admixture proportions (defined as the proportions of the genome that have ancestry from each subpopulation) vary between individuals (Hoggart *et al*. 2003).

Studies to determine association between a marker allele and the phenotype can take two forms. In one form, groups are distinguished on the basis of their divergent phenotypes (diseased vs. healthy; low vs. high trait value) and allele frequencies are compared across groups. Such studies are often referred to as case-control studies in the human genetics literature since they contrast disease-affected individuals (cases) with unaffected (control) individuals. The second type of study uses groups distinguished on the basis of their marker genotypes, and phenotypic means are compared across group.

Marker-trait association does not necessarily imply that markers showing a significant effect on the phenotype are linked to QTL. Rather, the marker-trait disequilibrium may exist in the absence of linkage, and instead may have arisen simply as a consequence of population structure. The relationship between the putative quantitative trait locus (QTL) and phenotype is the one of interest, but it can be confounded by other variables.

QTLs and individual admixture can be directly influenced by random variation due to meiosis. In addition, both the phenotype and measured admixture are potentially subject to measurement error. Furthermore, measured admixture is directly affected by individual admixture, which in turn is affected by individual ancestry. Naturally, the ancestry of the parents, represented by P1 and P2 affects individual ancestry. Individual ancestry can directly affect the putative QTL,

which in turn can affect the phenotype, so individual ancestry has an indirect affect on the phenotype via the putative QTL (Redden *et al*. 2006).

3. Assessment of Chickpea (*Cicer arientinum* L.) Germplasm Diversity for Crop Improvement

3.1 Introduction

Chickpea (*Cicer arietinum* L.) is the world's third most important pulse crop after bean and pea. It is an important pulse crop with a wide distribution across the tropics, subtropics and temperate regions (Singh, 1997). It accounts for about 15% (10.46 million tons) of the world's total pulse production (FAO, 2009).

The genus *Cicer* L. (Family Fabaceae) consists of 43 species including 34 perennial and eight wild species, and one domesticated chickpea, *Cicer arientinum* L. (van der Maesen *et al*. 2007). Chickpea is a self-pollinated crop, with $2n = 2x = 16$ chromosomes and a genome size of 740 Mb (Arumuganathan and Earle, 1991).

Cultivated chickpea is a small, herbaceous, annual shrub, showing considerable variation in form. Some types are semi-erect with a main stem and only a few branches, while others are semi-spreading with profuse branching. Normally the plants grow to a height of 18-24 inch (45-60 cm) and are frequently bluish to green color and covered with granular hairs. The tap-root is well developed, and can reach 30 cm or more in length (Kay, 1979).

Chickpea is originated in Asia and the eastern Mediterranean region. In ancient times cultivation quickly spread throughout the Mediterranean region and the South East Asian sub-continent, and gradually extended the drier parts of Africa, notably Ethiopia. Chickpeas were introduced successfully in to the New World, and have become an important crop in Mexico, Argentina and Chile. More recently it has been introduced in to Australia and Canada which becomes commercially important crop (Kay, 1979).

Chickpea is an edible legume with high in proteins and is one of the earliest cultivated grain legumes (Redden and Berger, 2007). In a report by Vavilov (1926), Southwest Asia and the Mediterranean were identified as the two primary centers of origin of chickpea, while Ethiopia as a secondary center of origin. Later, Singh (1997) reported that chickpea most probably originated in region of present day Southeastern Turkey and adjoining areas of Syria.

A detailed knowledge of the genetic relationships and diversity among accessions is an important factor for various aspects such as management of genetic resources, identification of duplicate accessions in the germplasm, selecting germplasm with desirable traits, in applied breeding programs and establishment of core collections (Dwevedi and Lal, 2009).

Accurate assessment of the levels and patterns of genetic diversity can also be invaluable in crop breeding for diverse applications including analysis of genetic variability in cultivars (Smith, 1984), identifying diverse parental combinations to create segregating progenies with a maximum genetic variability for further selection (Barret and Kidwell, 1998) and introgressing desirable genes from diverse germplasm in to the available genetic base (Thomson *et al*. 1998).

Chickpea has high variation for various qualitative and quantitative traits i.e. grain color and shape, color of flower, podding, plant height, yield, seed coat color, earliness and drought tolerance, insect pests resistance, like any other crop of different ecological zones, that can help breeders to release better and superior lines and varieties (Dasgupta *et al*. 1987; Singh, 1997).

To utilize properly the chickpea diversity present in the field and gene bank, there must be proper characterization and evaluation of the collected germplasm using multivariate analysis (cluster and principal component analysis).

Cluster analysis encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. A general question facing researchers in many areas of

inquiry is how to organize observed data into meaningful structures, that is, to develop taxonomies. In other words cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. Given the above, cluster analysis can be used to discover structures in data without providing an explanation/interpretation. In other words, cluster analysis simply discovers structures in data without explaining why they exist.

But principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components.

So the present study was undertaken to characterize Ethiopian chickpea germplasms using phenotypic and genotypic variability for management and efficient utilization of germplasms.

3.2 Materials and Methods

Genetic Materials

A total of 1032 chickpea accessions obtained from Institute of Biodiversity and Conservation (IBC) and three released varieties were evaluated in two environments at Kobo - research site of Sirinka Agricultural Research Center, North Eastern Ethiopia. The genetic materials were planted in randomized complete block design with two replications at two environments (non drought stressed and drought stressed environments) at Kobo.

Genotypes were evaluated for eleven quantitative traits i.e. days to 50 % flowering, days to maturity, plant height, biomass per plot, grain yield per plot, grain yield per plant, seed number per plant, pod per plant, hundred seeds weight, harvest index and pod filling period in 2010 main cropping season.

The Study Site

This study was conducted in Kobo, North Eastern Ethiopia which is one of research site of Sirinka Agriculture Research center. Kobo is located in North Eastern part of the country with latitude 12^0 08" N and longitude 39^0 28" E, and an altitude of 1470 meter above sea level (masl). It has 31^0C average annual temperatures with a range of 14 ^{0}C - 35 ^{0}C and 950 mm annual rainfall in the cropping season. The study site is characterized as a black soil that represents terminal drought stressed chickpea growing area in Ethiopia (Kobo metro station, 2010).

Cultural Operation

Germplasms were raised by direct seeding in the main field at a spacing of 30 cm between rows and 10 cm between plants. Each germplasm in each replication was grown in 2 rows with 20 plants per row. Supplemental irrigation was done two times on the crops critical growth stage (50 % flowering and pod setting stage) for non drought stress environment. Since the research site is representative of drought prone area for the country, the natural environment is considered as drought stress for the second set of the experiment. Other cultural operations were done as per the recommendation to the area during the crop growth period.

Method of Sampling and Recording of Observations

For plant height and other observations which were taken per plant, data were recorded based on randomly selected five plants on each plot. Mean values of five plants were used for statistical analysis. For other plant characters, observations were recorded on plot basis. The characters observed for eliciting the information are described below.

Days to 50 Percent flowering (days)

The number of days taken from sowing to 50 percent flowering in all plot was recorded.

Days to Maturity (days)

The number of days from the sowing to the physiological maturity was recorded for each plot.

Plant Height (cm)

The height of the plant from the base to the tip of the longest plant part was measured and expressed in centimeter.

Number of Pods per Plant at Maturity

The numbers of pods (both productive and non-productive) were counted at the time of harvest.

Number of Seeds per Plant

The numbers of seeds present per plant were counted at harvest. Observations were taken from five plants and the mean used for statistical analysis for that accession.

Yield per Plant (gm)

The total weight of all the filled grains per plant was measured in grams and recorded.

100 Seeds Weight (gm)

The weight of one hundred randomly selected filled grains was recorded in grams for each accession.

Biomass Yield (gm)

The total weight of all the plant above ground per plot was measured in kilo gram and recorded.

Harvest Index

It was computed from the ratio of grain yield to total biomass yield per plot and recorded.

Grain Yield per Plot (gm)

The total weight of all the filled grains per plot was measured in grams and recorded.

Grain Filling Period (days)

The time taken from 50 % flowering to physiological maturity per plot was recorded.

Data Analysis

Phenotypic data were analyzed using different statistical softwares (SAS V. 9.1, Agrobase V.33, Genstat V.12 and DARwin 5.0). Analysis of variance and correlation analysis were done using phenotypic data of the two environments (drought stressed and without drought stressed). Principal component analysis was also done using the two environment phenotypic data. The principal components that contributed for total variance were identified.

The means of quantitative traits' data were used for clustering the genotypes and calculating the genetic distance between groups. Hierarchical clustering was employed to determine the genetic structure of germplasm collections.

Genotypic data were also analyzed using DARwin 5.0 and hierarchical clustering using Ward method was employed to depict the distinct clusters based on SNP marker data.

3.3 Results

The analysis of variance for evaluated agronomic traits revealed that highly significant difference was observed on biomass per plot and harvest index, yield per plot and yield per plant; and a significant difference was also observed for the other evaluated agronomic traits except 50 % flowering days, hundred seeds weight and plant height among the tested accessions at without drought stressed environment.

Highly significant difference also observed in biomass per plot, pod per plant and grain yield per. But other studied traits showed non- significant difference among accessions at drought stress environment (Table 3.1).

Table 3.1 Summary of analysis of variance of phenotypic data at two environments

S.O.V	D.F	**Mean of Square at drought stressed environment**										
		BM	DF	HI	HSW	DM	PFP	PHT	PPP	SPP	YLD	YPP
Block	1	1.684**	6050.493ns	0.12**	200.35**	33.161**	5187.79**	110.84ns	4760.058**	102.667ns	60330.964**	91.501**
Genotype	1034	0.098**	9.04ns	0.003**	1.41ns	4.57ns	9.03ns	27.57ns	382.33**	196.04ns	3053.51**	6.72ns
Residual	1034	0.06	9.968	0.002	1.365	4.774	8.938	28.739	232.085	185.216	2459	6.701
CV %		18.94	6.42	22.42	7.71	2.35	6.83	11.74	24.5	17.75	18.45	17.9
		Mean of Square at without drought stressed environment										
Block	1	2.09**	8.417ns	0.209**	0.017ns	13.635ns	43.48*	561.43**	4.0ns	108.08ns	38174.61**	58.78**
Genotype	3034	0.25**	6.87ns	0.006**	2.043ns	8.298*	8.89*	19.752ns	145.981*	218.451ns	5637.425**	5.961**
	1034	0.213	7.327	0.005	1.942	7.456	8.326	20.619	110.237	210.468	4711.766	5.26
CV %		31.15	5.18	29.85	9.27	2.88	6.77	9.57	18.41	19.01	21.13	16.95

S.O. V = Source of variance, D.F = degree of freedom, BM = biomass per plot, DF = days to 50 % flowering, HI = harvest index, HSW = hundred seeds weight, DM = days to maturity, PFP = pod filling period, PHT = Plant height, PPP = pod per plant, SPP = seeds number per plant, YLD = yield per plot, YPP = yield per plant, * = significant at ($p< 0.05$) and ** = significant at ($p< 0.01$), ns = not significant

The combined analysis over location showed highly significant difference were observed for biomass per plot, yield per plot, pod per plant and harvest index for genotype and genotype x location interaction. But there was no significant difference between accessions for the other evaluated agronomic traits for genotype and genotypes x location interaction (Table 3.10).

Highly significant positive correlations were observed between yield and most of the evaluated agronomic traits except pod filling period and seeds number per plant. There was also highly significant positive correlation between harvest index and biomass, hundred seeds weight and seeds number per plant and non significant correlation was observed between days to 50 % flowering and biomass and plant height at drought stressed environments (Table 3.8).

The correlation analysis also revealed that highly significant positive correlation between yield and most of the studied traits except days to maturity and plant height, and harvest index showed highly significant positive correlation with most of the traits except days to maturity and plant height. So the correlation analysis revealed the most important traits that correlated and contributed positively for grain yield at both environments (Table 3.9).

Principal component analysis at two environments grouped the evaluated agronomic traits in to 11components which accounted for the entire (100 %) of the variability. It also showed that the first five vectors which have more than one eigen values, explained about 75 % of the total variance and 76.6 % of the total variance at drought stressed and without drought stress environments, respectively.

The first principal component axis explained the variation based on harvest index, hundred seeds weight, yield per plot and yield per plant. High loading vector were recorded for biomass per plot, pod filling period, pod per plant and yield per plant for the second PCA while the third eigen vector explained accessions based on biomass per plot, days to 50 % flowering and days to maturity. The fourth eigen vector explained by days to maturity, pod filling period and plant height indicating that traits are positively correlated with the total variance at drought stressed environment (Table 3.10 and 3.11).

At without drought stressed environment, the first principal component explained the variation based on biomass, pod per plant, seed number per plant, yield per plot and yield per plant. In the second component, harvest index and yield per plot were contributing high loading vector and the third eigen vector explained accessions based on days to maturity and pod filling period. So the traits showing higher positive loading value has strong correlation with the total variance since total variance is the linear combination of these values.

Hierarchical cluster analysis using Ward method revealed a dendrogram depicting the morphological relatedness and difference of chickpea accessions. This clustering algorithm successfully classified the total 1035 chickpea accessions and varieties in to nine morphologically distinct clusters for non drought stressed environment data and six morphologically distinct clusters for drought stressed environment data (Fig. 3.1 and 3.2).

In the first environment (non drought stress environment), cluster one contained 2 distinct accessions, cluster two contained 8 accessions, cluster three contained 57 accessions, cluster four contained 92 accessions, cluster five contained 91 accessions, cluster six contained 85 accessions, cluster seven contained 198 accessions, cluster eight contained 210 accessions and cluster nine contained 292 accessions.

At drought stress environment, cluster one contained 50 accessions, cluster two contained 229 accessions, cluster three contained 2 accessions, cluster four contained 242 accessions, cluster five contained 202 accessions and cluster six contained 309 accessions.

Cluster distance for each group and between the clusters was calculated for both environment data and the result revealed that higher distance was observed between cluster nine and cluster one (21.205) at non drought stress environment, and higher distance was also observed between cluster two and cluster four (12.984).

So the agronomic data based clustering showed that cluster nine at non stressed environment and cluster two at drought stress environment showed higher genetic diversity than other cluster (Table 9 and 10).

The Shannon-Weaver diversity index analysis for the entire collection revealed high genetic diversity for all the evaluated agronomic traits at both drought stressed and without drought stress environments. Comparatively, accessions showed better genetic diversity at with drought stress environment than without drought stress environment (Table 5.17).

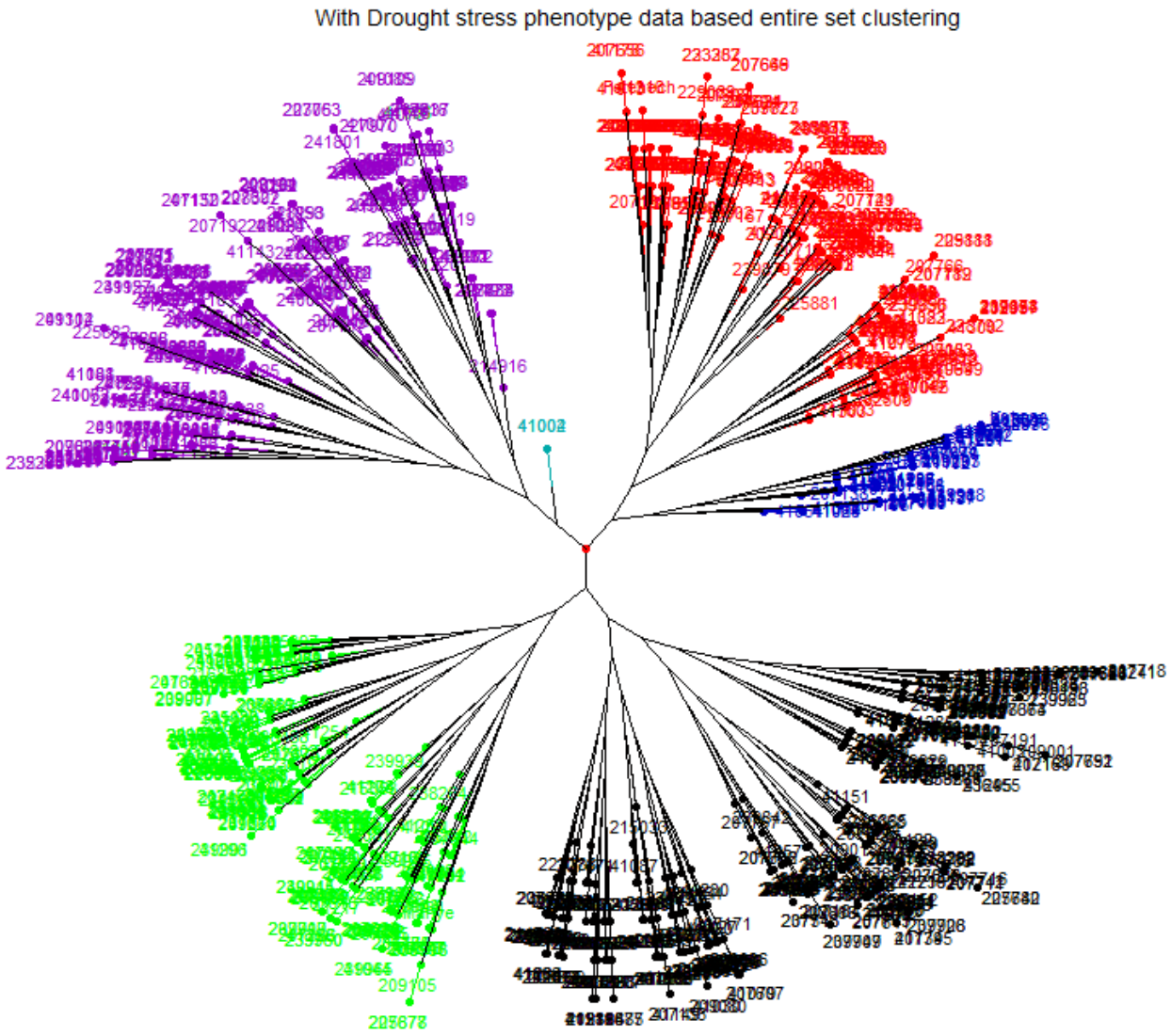

Fig.3.1 Schematic representation of entire germplasm collection based on phenotypic data clustering at without drought stressed environment (blue for cluster 1, red for cluster 2, light blue for cluster 3, rose for cluster 4, green for cluster 5 and black for cluster 6)

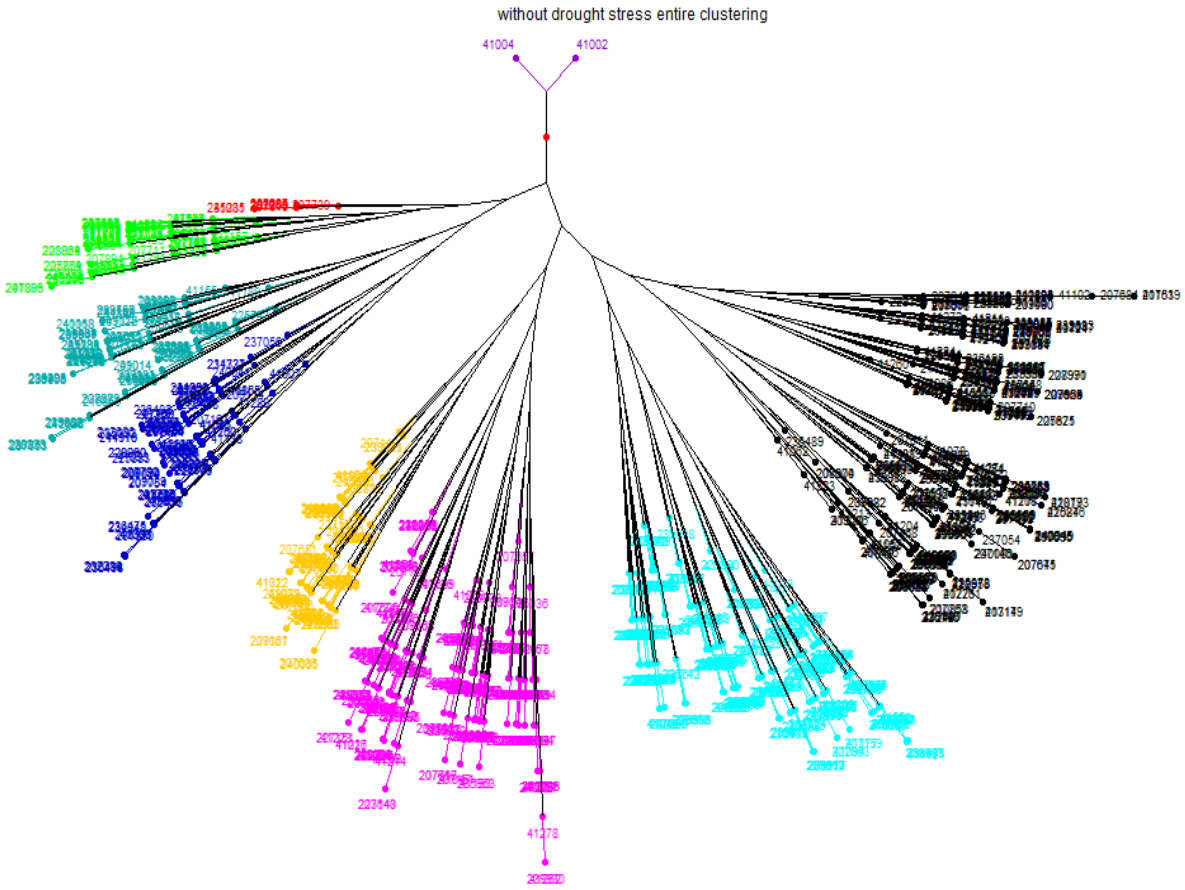

Fig.3.2. Schematic representation of entire germplasm collection based on phenotypic data clustering at drought stressed environment (rose for cluster 1, red for cluster 2, green for cluster 3, light blue for cluster 4, blue for cluster 5, yellow for cluster 6, light rose for cluster 7, very light blue for cluster 8 & black for cluster 9)

Hierarchical clustering analysis using Ward method for genotypic data also revealed a dendrogram that depicted the genotypic relatedness and difference of 1002 genotyped accessions and varieties. This clustering algorithm also clustered 1002 accessions and varieties in to six distinct major groups which are similar to morphological data based clustering. In the genotypic data clustering, cluster one contained 123 similar accessions, cluster two contained 14 similar accessions, cluster three contained 7 similar accessions, cluster four contained 532 similar accessions, cluster 5 contained 94 similar accessions and cluster six contained 240 similar accessions.

Even if the number of clusters was similar to morphological data clustering at drought stressed environment, the type and number of accessions present in each group for genotypic data based clustering was different from the morphological data based clustering (Table 3.4 & Fig. 3.3).

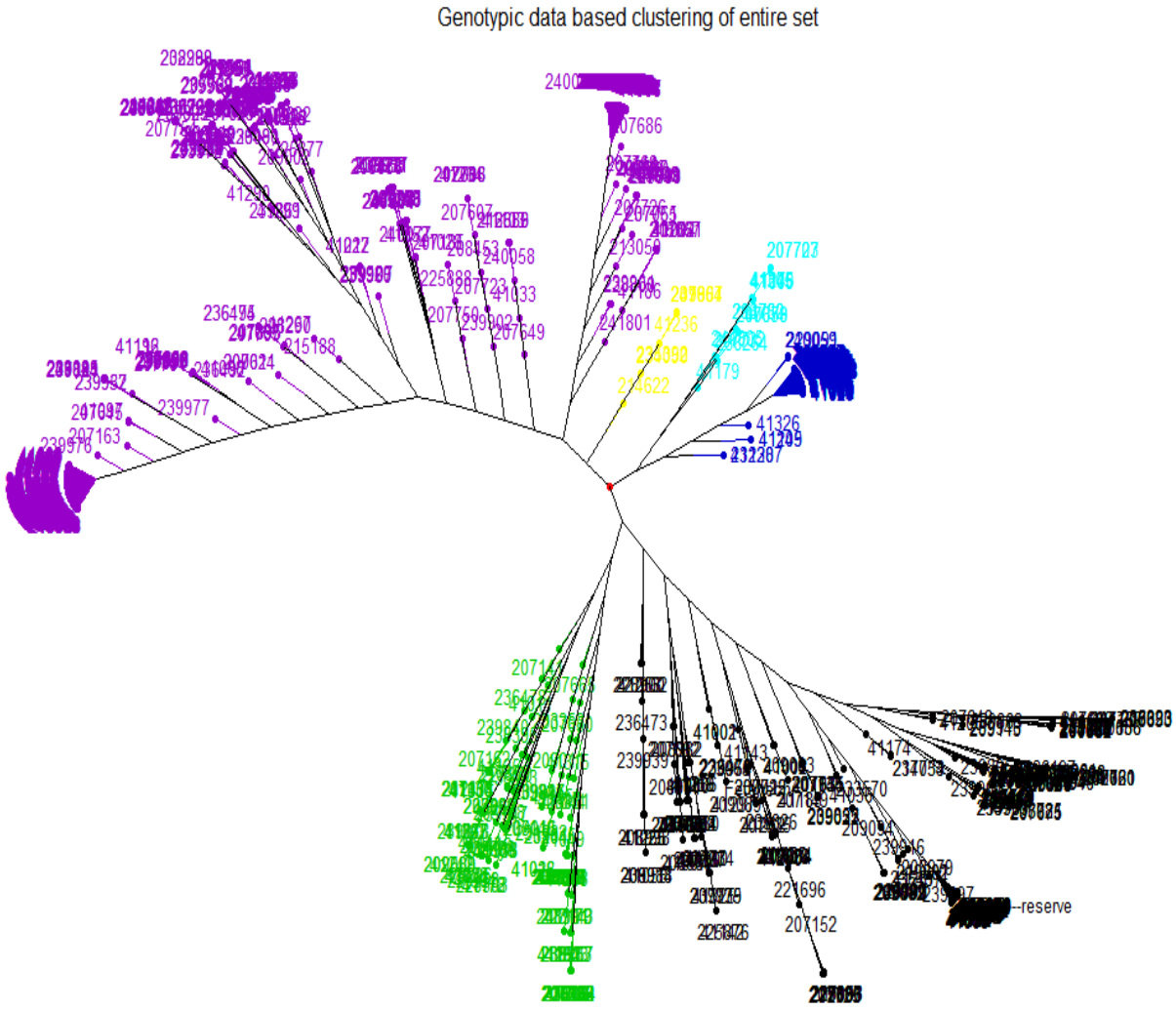

Fig. 3.3 Schematic representation of genotypic data based hierarchical clustering of entire collection (blue for cluster 1, light blue for cluster 2, yellow for cluster 3, rose for cluster 4, green for cluster 5 & black for cluster 6).

Table 3.2 Summary of combined over location analysis for all evaluated traits accessions

Sr no	Agronomic traits	Mean	R-Square	CV %	Significant level	
					Genotype	Loc x Geno.
1	50 % flowering days	50.	0.587	6.23	NS	NS
2	Maturity days	93.89	0.576	2.64	NS	NS
3	Plant height	46.53	0.503	10.73	NS	NS
4	Pod per plant	59.59	0.617	22.02	**	**
5	Seed per plant	76.48	0.511	18.39	NS	NS
6	Biomass per plot (kg)	1.38	0.583	26.78	**	**
7	Yield per plot(gm)	296.81	0.619	20.30	**	**
8	Hundred seeds weight	15.08	0.511	8.64	NS	NS
9	Yield per plant(gm)	13.99	0.533	17.47	NS	NS
10	Harvest index	0.225	0.554	27.40	**	**
11	Pod filling period	43.20	0.534	7.19	NS	NS

NS = not significant and ** = significant at p=0.001, Loc = location, Geno = Genotype

Table 3.3 Distribution of accessions in each cluster, clustering based on phenotypic data at without drought stressed environment

Cluster	List of accessions
I (2)	41002, 41004
II (8)	207687, 207709, 41084, 239253, 207739, 207616, 41281, 235035
III (57)	41027, 41134, 207706, 207188, 207647, 41195, 236490, 41284, 41088, 207750, 239844, 236881, 239876, 41029, 237055, 41280, 207724, 207163, 212687, 207729, 229703, 41045, 231329, 238265, 239878, 239980, 207723, 209001, 207644, 208453, 208984, 208829, 223064, 207147, 207742, 207766, 41165, 227148, 207711, 207894, 207764, 225879, 41139, 239892, 239879, 41294, 209009, 229956, 235722, 41290, 239863, 41196, 209012, 208364, 230253, 207895, 241803
IV (92)	41155, 209092, 209107, 233750, 41150, 209025, 41319, 234852, 41104, 41118, 240068, 41023, 41161, 41111, 41266, 228197, 236457, 209015, 41122, 236469, 241805, 209034, 239947, 41189, 209038, 209115, 212917, 207139, 219798, 41184, 235031, 207719, 207765, 240091, 209082, 41079, 41172, 207192, 207718, 41275, 208982, 209006, 207749, 41216, 239852, 207173, 214917, 228899, 207748, 228293, 209093, 235395, 236478, 239966, 209008, 239849, 239855, 239886, 239949, 207696, 239946, 41085, 41181, 207680, 209014, 41014, 225743, 236467, 41285, 223287, 227155, 228659, 232207, 207133, 229091, 41160, 235033, 41092, 207673, 228289, 214625, 236463, 216856, 219802, 225888, 41107, 228296, 239873, 212688, 207751, 237513, 227155
V (91)	237056, 214731, 231327, 41230, 41301, 234050, 236482, 212686, 235032, 41228, 207620, 207769, 41302,227160, 41170, 214916, 240041, 207663, 212477, 225887, 239850, 228199, 236477, 41020, 41154, 231328, 41003, 41286, 207135, 41296, 228290, 229960, 236488, 209033, 241126, 207630, 207753, 209104, 41093, 225883, 216855, 41163, 41130, 41146, 212916, 207720, 228792, 207167, 209010, 207170, 209094, 209114, 207617, 208980, 41312, 209003, 41006, 41068, 209026, 41041, 239836, 209089, 207754, 41007, 41289, 41100, 41250, 214728, 239914,228295, 207735, 41277, 207745, 41186, 214621, 236474, 41258, 236475,234048, 41138, 219804, 239877, 227149, 41185, 239865, 236476, 41300, 208991, 41083, 207734, 232288, 231330, 236473, 236194
VI (85)	207148, 41131, 41141, 240067, 41008, 207614, 228196, 207628, 236468, 240081, 239923, 209099, 208454, 207161, 235398, 239961, 240056, 41112, 207651, 41012, 41322, 236887, 207732, 41169, 239963, 207183, 41310, 227158, 207609, 41073, 229089, 41212, 239929, 239938, 41188, 41190, 207642, 233352, 41292, 221696, 240062, 41037, 240061, 207681, 41113, 207164, 208992, 215190, 207156, 209027, 227161, 207564, 209023, 41080, 41209, 41096, 209080, 239941, 41054, 41075, 41077, 207632, 41205, 239895, 207627, 41059, 209113, 209019, 41144, 214626, 214730, 240064, 41128, 41182, 41145, 202509, 41143, 236470, 41207, 41180, 41210, 228301, 229955, 207686, 240055
VII (198)	232206, 236455, 239928, 41171, 41259, 41265, 228198, 207150, 208988, 207615, 41234, 207626,

	207700, 41137, 236458, 41076, 207690, 41149, 41246, 41232, 239936, 227153, 239904, 235721, 41233, 41223, 227151, 41283, 207654, 212689, 41222, 233570, 209002, 239979, 207716, 207760, 240076, 41081, 41090, 207646, 209091, 207636, 41227, 41016, 41287, 207195, 236483, 207613, 41099, 207727, 228288, 228292, 207171, 41086, 207633, 207653, 209108, 41252, 41187, 236491, 209076, 41194, 228300, 235394, 232286, 41201, 232287, 207699, 41126, 41109, 207661, 41028, 227152, 41269, 41282, 207664, 207698, 239897, 212474, 239885, 41214, 207640, 223143, 207679, 41273, 209081, 239841, 41305, 41069, 41072, 41005, 207705, 239842, 41253, 41162, 207181, 240043, 235825, 41044, 239915, 41018, 41127, 207667, 215032, 41026, 225872, 207668, 207717, 219800, 41176, 41173, 219797, 41243, 233572, 41325, 207893, 41306, 41257, 207179, 239916, 41047, 207151, 41056, 207691, 239898, 236459, 207144, 207714, 41106, 207177, 41202, 41120, 207761, 235826, 207767, 207160, 207607, 209112, 239862, 41168, 41229, 236882, 41193, 235963, 209007, 207142, 41082, 219801, 41288, 229962, 207657, 240084, 235396, 239888, 207141, 216854, 41001, 41236, 227162, 241128, 207659, 207134, 207152, 41221, 236465, 41097, 207155, 41206, 236479, 239874, 239896, 228900, 41261, 41057, 41225, 207629, 207652, 41191, 207191, 235036, 207157, 240078, 41048, 207649, 41178, 41314, 225884, 41133, 235397, 41238, 240057, 231454, 41064, 240085, 223288, 41278, 41262, 239930
VIII (212)	207759, 214729, 236461, 209016, 235034, 239927, 236464, 207741, 209029, 239971, 228200, 235393, 239993, 239968, 228291, 241801, 207677, 240048, 239978, 41158, 207728, 207730, 41055, 41129, 207166, 207896, 209109, 207733, 214622, 207704, 41043, 41254, 208994, 41105, 41115, 41208, 41066, 207650, 240090, 207658, 207655, 214734, 239840, 41237, 208985, 207669, 227154, 207746, 225873, 239853, 235392, 239847, 225742, 207666, 239972, 207671, 209096, 228303, 207685, 207154, 41142, 41192, 239908, 209017, 41114, 236885, 239919, 239954, 41175, 239937, 207623, 207563, 207701, 41034, 209083, 209103, 207639, 207683, 240069, 41244, 41239, 41263, 239889, 41132, 239894, 41316, 207713, 215290, 41117, 41074, 207165, 209011, 240044, 41313, 239845, 208997, 207743, 239950, 41315, 228299, 41256, 41164, 225880, 207689, 240073, 207712, 207186, 239921, 240066, 208983, 238267, 41198, 212589, 236481, 41276, 209013, 236472, 41136, 41293, 207703, 209084, 41279, 207162, 208999, 241804, 236193, 240065, 41125, 207631, 41147, 41157, 207725, 207610, 239857, 214624, 208977, 209111, Mariye, 239851, 239890, 223065, 225877, 212478, 212685, 207738, 207656, 207561, 212475, 209078, 41270, 207606, 41094, 222863, 41040, 239952, 41052, 241800, 236454, 208989, 215353, 208993, 239965, 207562, 41200, 207707, 231331, 41303, 41177, 207159, 41033, 41321, 227970, 41071, 209105, 207892, 41010, 240077, 41217, 239860, 41183, 240087, 41140, 207185, 225874, 240088, 214732, 225738, 207744, 215033, 41318, 41299, 207726, 236883, 237057, 41151, 231332, 209102, 238264, 41215, 207145, 228294, 223142, 223063, 239838, 41022, 41116, 208987, 41051, 236471, 208978, 229090, 228658
IX (291)	41062, 41323, 41320, 209036, 41042, 207146, 207174, 209032, 207622, 209098, 41197, 207175, 41167, 41103, 225886, 41248, 236196, 225740, 239900, 41213, 207637, 210859,41060, 207182, 41152, 207638, 209101, 41058, 208979, 207763, 210858, 236489, 207674, 208900, 207682, 239922, 207168, 207634, 207672, 209031, 41309, 209116, 41226, 207731, 216853, 225882, 41255, 225889, 41245, 41308, 41159, 239909, 41209, 235720, 239909, 41204, 235720, 239967, 219803, 240079, 41297, 229958, 225878, 41317, 207149, 41098, 207694, 207770, 239912, 41070, 239864, 41065, 208998, 209004, 41030, 41108, 41311, 239932, Kutaye, 239911, 41324, 207618, 41199, 207140, 240080, 41089, 207608, 41031, 207136, 207608, 41031, 207136, 207692, 239903, 239977, 41241, 240063, 239870, 41247, 41298, 239846, 237054, 207641, 207675, 41231, 240089, 209088,239982, 240059, 41148, 236197, 207715, 207752, 41220, 213224 , 41038, 41268, 236492, 207138, 207662, 236884, 225890, 241127, 209106, 239959, 41009, 240045, 240049, 207611, 41021, 41011, 41249, 215577, 238262, 239902, 41267, 239917, 41101, 41218, 41063, 215289, 209022, 209110, 209085, 41219, 209030, 233571, 225876, 41024, 41078, 41251, 41274, 209018, 239861, 233353, 204785, 41271, 215189, 41017, 238793, 207670, 213051, 41291, 41036, 213050, 239924, 209021, 228298, 207702, 240054, 240070, 208986, 236886, 236493, 207172, 41025, 41135, 41039, 41121, 239901, 205148, 41307, 41260, 41061, 215188, 41156, 207710, 207621, 225875, 214733, 207693, 239945, 239935, 236480, 41264, 41242, 207721, 41095, 41035, 41046, 233573, 41032, 41091, 41326, 239976, 41015, 207648, 41019, 207665, 229959, 207153, 207624, 229961, 41110, 207768, 225881, 239925, 207635, 239906, 41224, 41235, 240042, 207736, 239960, 239957, 41067, 236196, 228297, 208990, 227971, 209020, 41240, 236462, 212476, 207184, 239891, 41304, 207178, 227156, 41174, 207645, 234049, 207688, 227972, 41013, 240071, 239918, 239905, 240058, 41179, 209086, 215067, 230795, 41272, 41053, 209090, 207612, 240060, 239859, 41124, 207747, 41211, 209087, 41203,

	239893, 227150, 207676, 208981, 219799, 215667, 41295, 228201, 239939, 207176, 207180, 209000, 209028, Fetenech, 41123, 41166, 207643, 41119, 239907, 41087, 207143, 207625, 241802, 41049, 41102, 207684, 41153, 207619,

Number in the parentheses indicating number of genotypes in each cluster

Table 3.4 Distribution of accessions in each cluster which was done using phenotypic data from drought stressed environment

Cluster	List of accessions in each cluster
I (50)	41054, 41029, 41064, 41012, 41062, 207195, 207186, 207609, 207563, 41137, 41124, 41133, 41291, 239968, 207138, 41074, 207188, 207564, 41094, 41187, 41237, 4238, 41205, 41236, 41008, 41160, 41245, 41102, 219797, 41023, 209103, 41070, 41252, 207640, 240071, 207760, 209009, 41251, 41297, 41071, 41226, 240062, 41203, 239976, 231327, 41271, 41309, 41159, 207680, Kutaye,
II (229)	41303, 202509, 237055, 240042, 241803, 239979, 241126, 41069, 239949, 240088, 208989, 227151, 227972, 239918, 41043, 236493, 239838, 207650, 207668, 41117, 209003, 227153, 240041, 239924, 41235, 41300, 228792, 41299, 239957, 207154, 239878, 41192, 230253, 41276, 41073, 41275, 41141, 41247, 239952, 41086, 41082, 41323, 232287, 233570, 41105, 236464, 229955, 239888, 207690, 239896, 239950, 41199, 41306, 41222, 223143, 207719, 207182, 209029, 239894, 239849, 240045, 41015, 228196, 41240, 207766, 209111, 225888, 225881, 209014, 207174, 209027, 41304, 231329, 219798, 240080, 207687, 239922, 41272, 239971, 207645, 239954, 239912, 41179, 207712, 209012, 228303, 228201, 41185, 239900, 41045, 207631, 41294, 228199, 234852, 209000, 207729, 207141, 207148, 207659, 207761, 41061, 239898, 236459, 207720, 236882, 215667, 41144, 225738, 207717, 41093, 225880, 209087, 41145, 239889, 240043, 212686, 225884, 208992, 234049, 41119, 207561, 209098, 239850, 239879, 41118, 207146, 235720, 209025, 228292, 209013, 41270, 207741, 210858, 41148, 209015, 235031, 233572, 240087, 207167, 235825, 239978, 237054, 241800, 207713, 209033, 207721, 236883, 41284, 207723, 239877, 219802, 41190, 207140, 207732, 241804, 207623, 207743, 235032, 207684, 214621, 208982, 207694, 207646, 207669, 209020, 239946, 227148, 227149, 41034, 207709, 41039, 41167, 41308, 207181, 41195, 41063, 239982, 207655, 41155, 207635, 240054, 236482, 41158, 208987, 229089, 233352, 233287, 207652, 216856, 239914, 207705, 228658, 207710, 208981, 207702, 207688, 236886, 207663, 239876, 239923, 41138, 208990, 239890, 225876, 41024, 239895, 239977, 207629, 207661, 239836, 41250, 41301, 207653, 225877, 41316, Fetenech, 207162, 41135, 232207, 207662, 41274, 207643, 41305, 212688, 228298, 41013, 41178, 207656, 239855,
III (2)	41002, 41004
IV (242)	214916, 207734, 225883, 41193, 212477, 41081, 239972, 41031, 41212, 41019, 41173, 41293, 41114, 233573, 41146, 41033, 207637, 212916, 41223, 239857, 207691, 210859, 236491, 215190, 41006, 41176, 213050, 214729, 41078, 41085, 209109, 41220, 207676, 221696, 41207, 228290, 214917, 41037, 207613, 41150, 41017, 41007, 227970, 209028, 215189, 209107, 215067, 213224, 207161, 41318, 239859, 227154, 41169, 41263, 41213, 219803, 229959, 209090, 214622, 239936, 212478, 41047, 41182, 241801, 207753, 223063, 41112, 207681, 207170, 41315, 209017, 207892, 41052, 41058, 228293, 41221, 209089, 41233, 209101, 228200, 41099, 207152, 41095, 41266, 207657, 41228, 228289, 41060, 41020, 228294, 209081, 207627, 228302, 209016, 41096, 241128, 41049, 41201, 207608, 207157, 207606, 240065, 41156, 240089, 41131, 41225, 41055, 207626, 240067, 41143, 207192, 41132, 207150, 209112, 228295, 41307, 240060, 41215, 209093, 41186, 207160, 236463, 209088, 239861, 41288, 214733, 41011, 208453, 240057, 41183, 207715, 233571, 41262, 229091, 207163, 207620, 209104, 41282, 207179, 209083, 236476, 41253, 239862, 41206, 207647, 235393, 239841, 41097, 41229, 207893, 208988, 41018, 41259, 41157, 239927, 41025, 41022, 41075, 41286, 207176, 41184, 240066, 41028, 208986, 207748, 229703, 240078, 41027, 208998, 207658, 228296, 41057, 214730, 209096, 41065, 225882, 209114, 41302, 239938, 41123, 240069, 207634, 239840, 41277, 41014, 41088, 207699, 228288, 41224, 41188, 41001, 41170, 209007, 228300, 207628, 41103, 229962, 240073, 41121, 235034, 41261, 240063, 41072, 213051,

	236492, 236881, 240056, 207155, 207617, 207738, 208983, 207727, 235396, 41035, 209034, 41162, 41056, 41083, 236194, 41324, 41046, 41255, 227150, 41110, 207750, 41009, 207638, 209002, 207667, 41067, 207621, 227727, 235963, 207172, 41194, 41168, 227157, 240079, 237057, 235395, 228659, 232288,
V (202)	207636, 209116, 41152, 207769, 235397, 41292, 41280, 205148, 41180, 41227, 41089, 207133, 41101, 41243, 41264, 41322, 41098, 239864, 215289, 229961, 238900, 239930, 216853, 41317, 207649, 208979, 209113, 209006, 209022, 207153, 230795,240077, 209106, 209110, 207754, 209091, 239967, 207611, 236469, 239897, 41180, 235036, 41120, 207896, 41242, 239966, 209076, 223142, 236481, 207733, 239908, 41076, 207725, 214625, 240091, 207180, 235394, 207739, 41026, 41134, 236198, 41278, 240048, 208997, 228198, 227724, 239860, 234050, 228197, 225886, 233353, 235826, 236488, 236483, 239844, 225873, 225889, 41254,241802, 229090, 239865, 207674, 212475, 239893, 212476, 207168, 41010, 239892, 41166, 236197, 207701, 236471, 207624, 240068, 231331, 239935, 207685, 239928, 41197, 207651, 207714, 208999, 207562, 219801, 207610, 239874, 41260, 41296, 239901, 41106, 41109, 207724, 239847, 239939, 41174, 41208, 215353, 41314, 236454, 41279, 207642, 228297, 236887, 41107, 41164, 41100, 41090, 207142, 41036, 207744, 227158, 239961, 215290, 240044, 239915, 240081, 207183, 207630, 41080, 41313, 207178, 207144, 208454, 239909, 209102, 207707, 239919, 41249, 41310, 239902, 207618, 41298, 207166, 209038, 235721, 41077, 41217, 233750, 239960, 238264, 41258, 209108, 227162, 41091, 239945, 227160, 239947, 207648, 207615, 214624, 209010, 209026, 214731, 216854, 207752, 207693, 41032, 209086, 207165, 239965, 41044, 41256, 209008, 207143, 239904, 225742, 41181, 207134, 236474, 204785, 207644, 207654, 239941, 41161, 41202, Mariye, 41038, 239873, 207147, 208993, 239916, 209105, 207677, 225878
VI (309)	207670, 214728, 240049, 236489, 207164, 207747, 41189, 239863, 240084, 41273, 240059, 239852, 41066, 41234, 223288, 207726, 235398, 241805, 207765, 207770, 207612, 212474, 232286, 240090, 208829, 41005, 207671, 41319, 228301, 41232, 207675, 41116, 240061, 240076, 209078, 207689, 239253, 41128, 207151, 41111, 41289, 41041, 209021, 231332, 209115, 215188, 207139, 207135, 41311, 208980, 207698, 207763, 208994, 236473, 219799, 207177, 212685, 236477, 41087, 41290, 208991, 223064, 212589, 209084, 212917, 41200, 207185, 41092, 41115, 41125, 41053, 207672, 41267, 207619, 239891, 207735, 41108, 41149, 207136, 41312, 41281, 41321, 41283, 225890, 215577, 228793, 236479, 41204,207683, 41198, 225874, 209099, 208984, 41265, 208985, 232206, 209080, 41030, 41104, 231454, 41147, 240085, 41209, 209092, 41130, 41325, 236475, 41127, 207679, 207695, 207759, 207731, 41175, 207736, 41230, 207171, 209036, 41079, 207607, 207767, 207632, 207706, 236470, 241127, 41196, 239845, 41042, 234048, 219804, 207746, 237513, 239842, 41257, 41211, 207704, 207633, 231330, 41244, 225887, 207664, 207616, 222863, 236480, 207184, 207749, 239907, 41269, 207696, 227155, 41059, 235033, 41219, 225743, 228291, 236458, 216855, 207641, 207703, 231328, 235392, 229956, 41210, 215032, 41165, 41172,208900, 235722, 239905, 41139, 207745, 209018, 209031, 236468, 239886, 209094, 227971, 239851, 239911, 207895, 227152, 207728, 239906, 207625, 207614, 209011, 207156, 225875, 208977, 41239, 227161, 41113, 41231, 236884, 41246, 207639, 223065, 240058, 41177, 207149, 236885, 209023, 212689, 41151, 41122, 207665, 207711, 207742, 214732, 238262, 228299, 207716, 207682, 225740, 41320, 240055, 207145, 239903, 236457, 239937, 240064, 209019, 236472, 239959, 41129, 207622, 41153, 41051, 236478, 41218, 239929, 238265, 239870, 236465, 239917, 209032, 239932, 240070, 209085, 236455, 41295, 41048, 41171, 237056, 41021, 41191, 209030, 207686, 207894, 219800, 41154, 41003, 41216, 207159, 41285, 207730, 236467, 207700, 225872, 41068, 207191, 209001, 207692, 207751, 207666, 209082, 239885, 236462, 239921, 41248, 207175, 41214, 229960, 41268, 207173, 212687, 225879, 207673, 208364, 208978, 41136, 235035, 236461, 229958, 41287, 239925, 239963, 209004, 41040, 236490, 236193, 236196, 41163, 228899, 41326, 41084, 41016, 207768, 239846, 238267, 214734, 207764, 214626, 239980, 239853, 41241, 207718, 227156, 215033, 41142, 41140, 41126

Number in parentheses indicate number of accessions in each cluster

Table 3.5 Distribution of accessions in each cluster, clustered based on genotypic data

Cluster	List of accessions in each cluster
I (123)	41138, 232207, 41109, 41245, 41326, 41029, 41255, 41106, 41225, 207625, 41037, 41039, 41042, 4160, 41069, 41070, 41071, 41072, 41074, 41077, 41079, 41082, 41083, 41084, 41089, 41090, 41091, 41098, 41099, 41105, 41106, 41107, 41114, 41119, 41153, 41158, 41183, 41225, 41229, 41230, 41232, 41233, 41235, 41237, 41239, 41241, 41243, 41244, 41246, 41247, 41248, 41249, 41250, 41251, 41254, 41255, 41256, 41257, 41265, 41278, 41279, 41323, 41324, 204785, 207133, 207136, 207138, 207172, 207622, 207625, 207639, 207642, 207670, 207700, 207707, 207731, 207733, 207739, 207741, 207748, 207768, 208829, 209004, 209019, 209020, 209081, 209103, 212687, 215190, 223287, 225740, 227155, 227971, 228198, 228291, 228292, 228295, 228297, 228301, 228302, 229955, 229956, 233573, 236467, 236468, 236472, 236483, 236881, 239852, 239860, 239873, 239888, 239891, 239898, 239907, 239921, 239937, 240041, 240048, 240087, 241127, 229091, 240059
II (14)	41179, 238264, 207659, 240078, 207732, 214732, 233750, 241805, 41005, 41175, 41240, 41305, 207703, 207727
III (7)	214622, 234050, 235392, 41236, 207654, 239967, 241804
IV (532)	241801, 41186, 41014, 207651, 41202, 215067, 238300, 239965, 213050, 207761, 237055, 207726, 207561, 207618, 225890, 209012, 234049, 241800, 41273, 207689, 209007, 207168, 207744, 207686, 41008, 41017, 41021, 41022, 41054, 41058, 41066, 41075, 41086, 41087, 41096, 41101, 41112, 41120, 41127, 41131, 41140, 41145, 41147, 41148, 41149, 41150, 41169, 41171, 41176, 41184, 41188, 41207, 41209, 41271, 41274, 41280, 41282, 41301, 41309, 41321, 207144, 207156, 207165, 207176, 207182, 207183, 207184, 207192, 207613, 207615, 207620, 207633, 207637, 207638, 207652, 207669, 207676, 207682, 207691, 207702, 207709, 207711, 207713, 207743, 207745, 208977, 208993, 208997, 209025, 209026, 209032, 209033, 209076, 209085, 209091, 209092, 209102, 209106, 209108, 209111, 214624, 214729, 215353, 219797, 225874, 227149, 227150, 227157, 228200, 228298, 229962, 231327, 231329, 231331, 231454, 232286, 235035, 236458, 236463, 236479, 237056, 239851, 239886, 239892, 239908, 239941, 239982, 240065, 240069, 240091, 241802, 241803, 240062re, 207649, 41033, 240058, 41103, 212689, 216856, 239902, 207723, 208453, 207607, 41264, 207706, 207718, 212686, 207750, 225888, 41028, 207185, 41252, 240077, 41159, 240071, 207766, 235398, 41104, 41208, 41298, 235031, 239917, 207641, 41222, 207653, 240061, 208454, 209078, 209107, 239929, 239960, 41027, 41212, 225877, 207632, 41160, 41187, 41214, 210858, 214734, 239959, 208981, 239923, Kutaye, 214916, 207644, 209104, 227153, 239980, 209002, 41073, 41129, 41165, 41215, 207643, 207656, 207657, 207716, 208991, 209038, 209096, 209114, 225875, 225882, 225883, 235396, 236459, 239838, 239879, 41061, 223064, 227161, 239950, 41268, 209031, 229090, 236481, 41316, 207178, 207698, 207742, 209084, 219799, 219803, 223288, 233571, 240085, 241126, 207695, 207721, 207562, 235720, 239968, 208999, 232288, 41299, 239861, 239849, 235722, 238265, 208992, 207729, 239253, 41085, 41269, 207191, 41166, 207145, 207623, 239905, 41290, 41020, 41034, 41052, 41284, 41289, 209029, 212917, 239862, 239901, 233572, 207754, 41045, 41296, 207760, 228293, 239847, 239865, 239965, 240042, 41267, 228290, 215188, 41067, 207895, 207705, 236194, 236475, 41068, 236492, 41292, 41303, 207159, 209006, 214730, 230253, 235825, 236198, 239859, 207701, 41116, 41192, 239977, 239932, 239957, 41181, 228199, 239945, 207624, 239894, 41097, 207645, 207163, 239976, 41006, 41009, 41010, 41019, 41023, 41024, 41046, 41047, 41048, 41051, 41056, 41062, 41063,41076, 41088, 41092, 41094, 41095, 41106, 41121, 41123, 41124, 41125, 41126, 41132, 41133, 41139, 41144, 41146, 41151, 41154, 41155, 41156, 41157, 41161, 41163, 41164, 41167, 41168, 41180, 41194, 41197, 41213, 41216, 41219, 41221, 41223, 41224, 41226, 41227, 41228, 41231, 41238, 41258, 41261, 41263, 41270, 41272, 41275, 41276, 41277, 41281, 41295, 41306, 41310, 41311, 41313, 41319, 41320, 41322, 207139, 207142, 207143, 207146, 207147, 207150, 207157, 207160, 207164, 207166, 207174, 207175, 207177, 207180, 207186, 207563, 207564, 207608, 207612, 207617, 207619, 207626, 207627, 207628, 207629, 207634, 207647, 207666, 207667, 207668, 207671, 207673, 207677, 207680, 207690, 207699, 207704, 207712, 207717, 207719, 2-07724, 207736, 207747, 207751, 207753, 207759, 207894, 208976, 208983, 208986, 208990, 209000, 209001, 209008, 209009, 209013, 209034, 209067, 209105, 209110, 209112, 212475, 212476, 212477, 212916, 213051, 213224, 214621, 214626, 214917, 215033, 215189, 215289, 215290, 215570, 216855, 219800, 219804, 222863, 223063, 225873, 225880, 225884, 225886, 225889, 227160, 228196, 228197, 228289, 228294, 228299, 228899, 229089, 229959, 229961, 231328, 232287, 234048, 236454, 236461, 236465, 236469, 236474, 236477, 236480, 236482, 236488, 236489, 236493, 236883, 236884, 236886, 236887, 237057, 239836, 239842, 239846, 239850, 239857, 239864, 239874,

	239876, 239885, 239911, 239915, 239924, 239925, 239927, 239930, 239938, 239949, 239966, 239972, 239978, 240043, 240049, 240054, 240055, 240063, 240064, 240066, 240068, 240073, 240080, 241128, 41059, 41078, 41108, 41218, 208978, 209087, 207176, 235394, 207614, 209021, 228300, 239961, 225879, 41218,
V (94)	207141, 236478, 239840, 207162, 41136, 207153, 212474, 41031, 41308, 41011, 235395, 41135, 41302, 239878, 41065, 41217, 41242, 202509, 209080,210859, 207694, 233352, 239936, 227970, 41260, 41141, 41038, 41081, 228296, 236193, 205148, 209116, 41304, 229958, 239947, 239893, Mariye, 239918, 239935, 41189,41172, 41178, 228303, 207665, 233353, 239841, 239896, 209017, 207140, 209016, 207650, 209011, 41195, 231332, 41018, 41026, 41315, 41117, 209015, 239845, 41044, 225881, 41049, 41162, 41288, 41291, 41294, 41300, 41317, 207155, 207167, 207728, 207769, 228792, 235963, 239890, 240049, 41283, 208994, 41152, 41293, 41111, 207179, 223143, 41043, 239903, 238267, 41285, 41314, 207135, 207752, 214625, 223142, 229960, 236196, 236464
VI (240)	41253, 228900, 227162, 236473, 239939, 41025, 208998, 41266, 41055, 208988, 239914, 207692, 208984, 207663, 208900, 208982, 41191, 41312, 207170, 41307, 208980, 207173, 234852, 207611, 207674, 41030, 231330, 41015, 41130, 207687, 41016, 41259, 41286, 41110, 207734, 41325, 207735, 239979, 41142, 225876, 41001, 41002, 223065, 239954, 236476, 41297, 209089, 41143, Fetenech, 41262, 207635, 207715, 239922, 235826, 41200, 41205, 207616, 239928, 207684, 209014, 41032, 221696, 207152, 207725, 212685, 225887, 227156, 209093, 227151, 41004, 41012, 209003, 41185, 207149, 41193, 207134, 207746, 207655, 209027, 235032, 239855, 41036, 209094, 209082, 209099, 209101, 236882, 225742, 228793, 240090, 233570, 239916, 41287, 236457, 219801, 219802, 208979, 239897, 41003, 41035, 41080, 41128, 41170, 41173, 41182, 41190, 41196, 41198, 41199, 41203, 41204, 207195, 207606, 207664, 207672, 207679, 207681, 207683, 207993, 207696, 207710, 207714, 207749, 207763, 207764, 209010, 209018, 209028, 209030, 209036, 209086, 209088, 209109, 209115, 212688, 215032, 216854, 227152, 228201, 229703, 230795, 235033, 236470, 238262, 239844, 239853, 239877, 239919, 239971, 240088, 240089, 207610re, 41174, 214733, 237054, 239904, 214731, 235397, 239909, 41134, 41177, 41201, 41211,41220, 207151, 207171, 212589, 228288, 236462, 236491, 237513, 239889, 239912, 207730, 207675, 208985, 235721, 239895, 41093, 41102, 207161, 207646, 207685, 228658, 41137, 207765, 207770, 209090, 225872, 239900, 235034, 207181, 207621, 209083, 225878, 227154, 235036, 240044, 209098, 236471, 41007, 239906, 41040, 41041, 236197, 41210, 207896, 208364, 208989, 232206, 41064, 41216, 207631, 207720, 240060, 41206, 209113, 239946, 207648, 239863, 41057, 207658, 207661, 207662, 207688, 214728, 215667, 219798, 240067, 225743, 41053, 207188, 207636, 41113, 207609, 41115, 207767, 207630, 216853, 209023, 235393,

Number in parentheses indicate accession numbers in each cluster group

Table 3.6 Cluster distance of each cluster calculated using phenotypic data at without drought stressed environment

cluster	I	II	III	IV	V	VI	VII	VIII	IX
I	**7.61332**								
II	8.22433	**4.59553**							
III	8.07617	4.72705	**3.90345**						
IV	8.36763	4.81132	4.04397	**4.11871**					
V	8.16309	5.15393	4.34885	4.7331	**5.35895**				
VI	7.99275	4.84567	4.03103	4.33501	5.76229	**3.5662**			
VII	8.06716	4.83932	4.06479	4.27396	5.61574	3.67037	**3.48732**		
VIII	8.1488	4.72442	3.96689	4.23918	5.63747	3.67882	3.54785	**3.2777**	
IX	21.20526	16.34097	16.68781	15.70756	19.34972	16.48047	16.33374	16.2286	**12.49802**

Bold diagonal values indicate within cluster distance

Table 3.7 Cluster distance of each cluster calculated from data of drought stressed environment

cluster	I	II	III	IV	V	VI
I	**4.88469**					
II	7.46378	**9.72543**				
III	5.33811	12.04254	**3.4984**			
IV	5.60378	12.98416	3.66422	**3.45562**		
V	5.59015	12.87018	3.80074	4.90656	**3.70912**	
VI	5.50107	12.24138	3.61358	4.91315	3.87255	**3.37642**

Bold diagonal values indicate within cluster distance

Table 3.8 Pearson's correlation of evaluated agronomic traits at without drought stress environment

Trait	DF	DM	BM	YLD	HSW	HI	PFP	PPP	YPP	PHT
DM	0.4248									
BM	0.0318	0.0358								
YLD	0.0525*	0.0479**	0.3068**							
HSW	0.0494*	0.0638**	0.2902**	0.5193**						
HI	0.0284	0.0187	0.6885**	0.4242**	0.0959**					
PFP	0.5013**	0.5703**	0.0054	0.934	0.4617	0.0079				
PPP	0.0529*	0.0425	0.2442**	0.2518**	0.258	0.0518*	0.0074			
YPP	0.0615**	0.0490*	0.1592**	0.2149**	0.2260**	0.0044	0.009	0.4228**		
PHT	0.03	0.1148**	0.1022**	0.1772**	0.1591**	0.282	0.0825**	0.0731**	0.1034**	
SPP	0.0629**	0.0589**	0.2635	0.5167	0.3849	0.1273**	0.0008	0.3807	0.2697	0.1017**

BM = biomass per plot, DF = days to 50 % flowering, HI = harvest index, HSW = hundred seeds weight, DM = days to maturity, PFP = pod filling period, PHT = Plant height, PPP = pod per plant, SPP = seeds number per plant, YLD = yield per plot, YPP = yield per plant, * = significant at ($P< 0.05$) and ** = significant at ($P< 0.01$)

Table 3.9 Mean of evaluated traits in each cluster without drought stressed environment

Trait/ mean	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
MB	1.67	1.72	1.49	1.52	1.53	1.43	1.46	1.52	1.45
DF	52.50	52.75	52.41	51.97	52.05	52.48	52.05	52.19	52.35
HI	0.27	0.21	0.24	0.24	0.24	0.24	0.24	0.23	0.24
DM	96.50	94.44	94.79	94.84	95.00	95.18	94.83	94.80	94.85
PFP	44.00	41.69	42.38	42.88	42.95	42.70	42.77	42.61	42.50
PHT	48.75	46.13	48.31	46.66	47.43	46.78	47.47	47.65	47.49
PPP	75.75	56.69	57.28	58.61	58.41	55.31	55.92	58.32	56.41
SPP	83.50	81.25	79.12	77.55	78.04	73.19	75.18	76.87	75.96
HSW	16.25	15.06	15.11	15.20	15.03	15.06	14.94	15.04	14.99
YLD	419.75	324.35	334.07	331.70	331.32	315.42	319.83	327.55	322.69
YPP	18.48	13.04	14.07	13.48	13.40	13.25	13.55	13.65	13.44

BM= biomass per plot (kg), DF = 50 % flowering days, HI = harvest index, DM = maturity days, PFP = pod filling period, PHT = plant height, PPP= pod per plant, SPP = seed number per plant, HSW = hundred seeds weight (gm), YLD = yield per plot (gm), YPP = yield per plant (gm)

Table 3.10 Mean of evaluated traits in each cluster at drought stressed environment

Trait/ mean	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
MB	1.32	1.28	1.77	1.33	1.27	1.30
DF	49.02	49.03	48.00	49.03	49.24	49.27
HI	0.22	0.21	0.20	0.21	0.22	0.22
DM	93.29	93.11	95.25	92.84	92.76	92.84
PFP	43.85	43.96	43.25	43.55	43.91	43.70
PHT	46.63	45.91	42.50	45.35	45.65	45.58
PPP	62.77	60.40	83.25	63.65	61.69	62.41
SPP	74.68	76.47	81.75	77.94	76.04	76.53
HSW	14.95	15.12	15.50	15.16	15.09	15.24
YLD	266.86	263.89	481.25	274.79	263.35	270.37
YPP	14.78	14.13	17.94	14.66	14.33	14.57

BM= biomass per plot (kg), DF = Days 50 % flowering, HI = harvest index, DM = days maturity, PFP = pod filling period, PHT = plant height, PPP= pod per plant, SPP = seed number per plant, HSW = hundred seeds weight (gm), YLD = yield per plot (gm), YPP = yield per plant (gm)

Table 3.11 Pearson's correlation of evaluated agronomic traits at drought stressed environment

Trait	DF	DM	BM	YLD	HSW	HI	PFP	PPP	YPP	PHT
DM	0.3686**									
BM	-0.0571**	-0.008ns								
YLD	0.0878**	-0.027ns	0.2544**							
HSW	0.1494**	0.0545*	0.1035**	0.5977**						
HI	0.1026**	-0.019ns	-0.5907**	0.6045**	0.3977**					
PFP	-0.8046**	0.2554**	0.0542*	-0.108**	-0.121**	-0.119**				
PPP	-0.0711**	-0.0174ns	0.2216**	0.1672**	0.0992**	-0.023ns	0.0629**			
YPP	-0.0176ns	-0.0038ns	0.1819**	0.2626**	0.1628**	0.0784**	0.0159ns	0.4191**		
PHT	0.1074**	0.1782**	-0.0001ns	-0.0145ns	0.0497*	-0.021ns	0.0021ns	0.012ns	-0.0044ns	
SPP	0.0816**	0.0577**	0.0749**	0.2682**	0.2599**	0.1705**	-0.048*	0.0784**	0.0599**	0.0459*

BM = biomass per plot, DF = days to 50 % flowering, HI = harvest index, HSW = hundred seeds weight, DM = days to maturity, PFP = pod filling period, PHT = Plant height, PPP = pod per plant, SPP = seeds number per plant, YLD = yield per plot, YPP = yield per plant, * = significant at (P< 0.05) and ** = significant at (P< 0.01)

Table 3.12 Eigenvectors and the percentage variations for the traits revealed by principal component analysis(PCA) at drought stressed environment

	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5
Eigenvector	2.441	1.892	1.572	1.306	1.053
Total variance %	22.189	39.3907	53.679	65.548	75.116
Eigenvectors (loading) for the traits					
Biomass per plot	-0.0122	0.4185	0.5117	-0.0828	-0.4121
Days to 50 % flowering	0.2832	-0.4731	0.4637	0.0338	0.117
Harvest index	0.4415	-0.1523	-0.5062	0.0384	0.1898
Hundred seeds weight (gm)	0.4782	0.1157	-0.0499	0.0812	-0.2444
Days to maturity	0.0414	-0.0855	0.1966	0.7307	0.1129
Pod filling period (days)	-0.2682	0.4374	-0.3566	0.4316	-0.0496
Plant height(cm)	0.0401	-0.0582	0.1679	0.4755	0.0945
Pod per plant	0.1377	0.4129	0.1937	-0.0682	0.4863
Seeds number per plant	0.2801	0.0894	0.0182	0.1586	-0.3875
Yield per plot (gm)	0.5254	0.2116	-0.1002	-0.0326	-0.1887
Yield per plant (gm)	0.2173	0.3732	0.1502	-0.0821	0.5239

Table 3.13 Eigenvectors and the percentage variations for the traits revealed by principal component analysis (PCA) at without drought stress environment

	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5
Eigenvector	2.6836	1.7206	1.6017	1.3928	1.0291
Total variance %	24.3963	40.0384	54.5996	67.2612	76.6162
Eigen vectors (loading) for the traits					
Biomass per plot	0.3123	-0.5796	-0.0206	-0.0141	-0.2759
Days to 50 % flowering	0.1008	0.0607	-0.2539	0.7858	-0.035
Harvest index	0.0398	0.7486	-0.0017	-0.0549	0.0413
Days to maturity	0.1153	0.0364	0.5602	0.5607	0.0509
Pod filling period (days)	0.0188	-0.0203	0.7658	-0.1771	0.0803
Plant height(cm)	0.1825	0.0287	0.1611	0.0606	-0.3549
Pod per plant	0.3761	-0.1046	-0.0593	-0.0395	0.5217
Seeds number per plant	0.4472	0.0877	-0.0459	-0.0832	0.0092
Hundred seeds weight (gm)	0.4268	0.0672	-0.0082	-0.0849	-0.275
Yield per plot (gm)	0.4598	0.2683	-0.032	-0.119	-0.2905
Yield per plant (gm)	0.3297	-0.0481	-0.0463	-0.0058	0.5947

3.4 Discussion

Plant genetic resources or germplasms are the most valuable, essential, and basic raw materials for crop improvement programmes to meet the demands of increasing populations. Vavilov (1926) was the first geneticist to realize the essential need for a broader genetic base for crop improvement by collecting germplasm of crops and their wild relatives globally.

Genetic diversity is the amount of genotypic (on the DNA level) variability present in a group of individuals or genotypes. Genetic diversity gives species the ability to adapt to changing environments, including new pest, disease and new climatic conditions, such as global warming (Parmesan and Yohe, 2003). Genetic diversity commonly is measured by genetic distance (GD) or genetic similarity (GS = GD-1), both of each imply that there are either differences or similarities at the genetic level (Weir, 1990).

Characterizing and assessing genetic variability of germplasm collection and making accessible to the breeding program is allowing geneticist and breeders to employ improved strategies to develop more efficient selection methods and genetic populations (Nyquist, 1991).

The assessment of genetic diversity is important not only for crop improvement but also for efficient management and conservation of germplasm resources and identification of duplicate accessions in the gene bank collections.

Even if chickpea is one of the crops with narrow genetic diversity in grain legumes (Ahmed and Slinkard, 1992), the present study revealed that Ethiopian chickpea germplasm collection showed high morphological and molecular genetic diversity. The performance of chickpea accessions at contrasting environment revealed significant morphological differences. Most of the accessions showed yield stability and better harvest index across the tested environments (without drought stress and with drought). Geographically, Accessions from Amara region revealed better performance and morphological and molecular diversity.

Cluster analysis of accessions showed that accessions collected in the earlier time showed better diversity than recent collections and better diversity observed at drought stressed environment than without drought stress environment. This showed that the genetic diversity eroded and narrowing through time and the finding also revealed that drought stressed environment is the better environment for unlocking and expressing the hidden genes and to assess genetic diversity. Similar findings were reported on erosion of genetic diversity and expression of genetic diversity in different crops (Bayush and Berg, 2007)

Since significant associations observed between grain yield per plot and other yield component traits (yield per plant, biomass per plot, pod per plot and hundred seeds weight), simultaneous improvement and selection of yield component traits should be feasible. The principal component

analysis result revealed that five main components contributed 75-76.6 % of total variability at drought stressed and without drought stress environments, respectively. Similar findings reported in different crops.

4. Population Structure and Association Analysis of Chickpea Germplasm

4.1. Introduction

Cultivated chickpea (*Cicer arietinum* L.) is a self pollinated, diploid (2n = 2x = 16) annual pulse crop with a relatively small genome size of 750 Mbp (Arumuganathan and Earle, 1991). It is an important legume crop of the semi-arid tropics and the West Asia, North and East Africa regions. It ranks third among pulses, fifth among grain legumes, and 15th among grain crops of the world in area coverage. It is cultivated on 11.55 million ha in the world with 10.46 million tons of production per annum (FAO, 2009). About 97% of the chickpea area is in developing countries, where it is largely grown under marginal and moisture stress condition.

Ethiopia is the largest producer of chickpea in Africa accounting for about 63 % of the continent's production. It is the second most important pulse crop after faba bean in production and third in area coverage after faba bean and field pea that contributes 16 % of the total pulse production in the country. The total annual chickpea production is estimated about 233 thousand tones and the national average chickpea yield is 1.3 ton/ha (FAO, 2009). The crop also fetches high domestic and export market potential and earning about 22 million US dollar annually (ECuA, 2008).

Chickpea is valued for its nutritive seeds with high protein content. Seeds are eaten fresh as green vegetables, fried, roasted and boiled, as snack food, sweet and condiments and flour as soup. It is grown in Ethiopia with 95 % *desi* and 5% *kabuli* type with different values.

Since Ethiopia has diverse agro-ecological systems, it is the center of diversity for desi type of chickpea with huge germplasm diversity present in the farmers' field and gene bank of Ethiopia. Characterizing and developing information about these germplasm collections is a great importance for both the conservation and utilization of genetic resources present in gene bank.

Because of the diverse nature of gene bank germplasm materials (landraces, released varieties and wild and weedy relatives from different areas of origin), they provide all relevant allelic diversity necessary for plant improvement. So, naturally diverse germplasm are suitable for trait - marker association studies and developing elite lines with special merits like drought or disease resistance (D'hoop *et al*. 2010).

However the large number of accessions accumulated in gene banks reduces the efficiency and effectiveness with which these genetic resources can be exploited. So determination of the genetic structure of heterogeneous germplasm collections is an essential component in the utilization, conservation or sampling of core collections.

In addition, it may be necessary to associate accessions in the core collection with the entire collection which is based on the group structure.

The determination of genetic structures of germplasm collections is also an important aspect of association studies (Shriner *et al*. 2007). General agreement exists among researchers that incorporating population structure in to statistical models used in association mapping is necessary to avoid false positives (Pritchard *et al*. 2000b; Flint-Garcia *et al*. 2003; Zhu *et al*. 2008). The general model for association mapping can be written as "phenotype = marker + genotype + error", and test for a marker effect is equivalent to testing to QTL. Typically, genotype is a random factor whose effects are structured by kinship or population structure. The relationship between phenotype and marker can be tested within different groups or genetic groups which can be used as an extra factor or as a covariate in modeling the relationship (Thomsberry *et al*. 2001; Wilson *et al*. 2004). Yu *et al*. (2006) further described by introducing a mixed model approach which incorporates both population structure (Q) and kinship (K) in modeling the relationship between phenotype and marker.

In recent years, many new methods have been developed especially for studying structure in natural populations using molecular markers, e.g. STRUCTURE (Pritchard *et al*. 2000).

Association mapping or linkage disequilibrium mapping is a new method that exploits the variation in a collection of genetically diverse materials (composed of unrelated individuals or unknown pedigrees) to uncover a significant association between a trait and a gene or a molecular marker on the basis of linkage disequilibrium. Association mapping offers the advantage that historical and evolutionary recombination can be exploited at the population level and all natural genetic diversity (larger number of alleles studied) can be used in order to obtain a high – resolution map. Moreover, no pedigree or cross is required, making it easier to produce the data (Aranzana *et al*. 2005).

So the objective of this study was to identify trait marker association and quantification of linkage disequilibrium in a structured natural chickpea germplasm using SNP markers.

4.2. Materials and Method

Plant materials

Nine hundred and ninety nine Ethiopian chickpea accessions and three release varieties were used to study the marker-trait association (Table 4.1). These accessions included the whole collection of the country present in the gene bank of Ethiopia. These accessions were collected at major chickpea growing states- Amara (453), Oromiay (285), SNNP (43), Tigray and Eritra (65), Unknown (150), Somali (3), Benishangul and Gumz (2) and Gambel (1). Accessions were collected from 1300 meter above sea level (masl) to 3200 masl since 1970 till 2000. The varieties were chosen based on its geographical history (representing major growing state) and drought stress reaction (two drought resistant and one drought susceptible).

Phenotypic evaluation

Accessions were evaluated at one of chickpea growing areas (Sirinka Agricultural Research Center - Kobo research site) that represents drought stress environment in North eastern Ethiopia. A total of 1032 accessions and 3 released varieties were planted at Kobo in randomized complete block design in two replications with two sets. Experiment set one was conducted at non drought stressed environment (created by applying supplemental irrigation at 50 % flowering and pod setting). The second set of experiment was evaluated at drought stress environment (natural environment that represents terminal drought in Ethiopia) in the same place. Accessions were planted in two rows spaced 30 cm between rows and 10 cm between plants. Accessions were evaluated for different agronomic traits such as 50 % flowering days, Maturity days, plant height (cm), biomass per plot (kg), grain yield per plot(gm), grain yield per plant(gm), hundred seeds weight (gm), seed number per plant, pod number per plant, harvest index and pod filling period at two environments.

DNA Extraction and High-throughput Genotyping

1002 Ethiopian chickpea germplasms (999 accessions and 3 varieties) were sown in green house at ICRISAT- Patancheru campus. Seedlings were raised in pots up to leaf sample reach for DNA extraction stage. Leaf samples at the age of 15-20 days were used for DNA extraction. DNA was isolated as per high-throughput mini DNA extraction protocol of ICRISAT. Quality and quantity of DNA was checked using agaros gel electrophoresis and the concentration was normalized. 115 informative, polymorphic and mapped (ICC4958 X PI 489777 interspecific cross of chickpea, unpublished) SNP markers, which cover the whole genome of chickpea with uniform distribution in each chromosome, were selected and used for genotyping 1002 chickpea germplasms.

SNP genotyping was done at KBioscience- UK, with relatively new, high through-put genotyping procedure, the KBioscience Competitive Allele- Specific Polymerase chain reaction (KASPar) assay (Orru *et al*. 2009). From 115 SNP markers, 111 SNP markers produced meaningful data with allele calling success rate of 99 %. So a total of 11078 allele calls in 112776 well were done. Polymorphic SNP markers were viewed with SNPviewer2 software graphically with KASPer validation kit that contains three separate tubes of flours representing the three observable genotyping groups.

Data analysis

We used the model based STRUCTURE 2.3.3 software to evaluate the optimal number of clusters and to assign each individual to a corresponding subgroup without using the predefined information. The software STRUCTURE adopts a Bayesian cluster approach under the main assumptions of HWE within populations and complete linkage equilibrium between loci within populations for markers not in admixture linkage disequilibrium.

STRUCTURE was run for the number of fixed subgroups K from 1 to 10, and five runs were performed for each K with a burn in of 100,000 followed by 100,000 cycles of replication for the actual analysis which produced Q matrix, assuming admixture of populations. As the STRUCTURE software overestimates the number of subgroups for the accessions, it is difficult to choose the ‘‘correct’’ K from the Ln probability of data, Ln P (D). Thus, the correct K value was decided based on Delta K value estimated by STRUCTUREHARVEST software and DARwin 5.0 cluster analysis.

Trait –marker association analysis was done using, “Trait Analysis by Association, Evolution and Linkage” (TASSEL) 3.0 standalone software (Churchill *et al.* 2004). We have employed general linear model (GLM) and mixed linear model (MLM) to increase accuracy and resolution of

association analysis. The relative kinship coefficients (K- matrix) among all pairs of accessions were calculated using 55 polymorphic SNP marker data with this program.

Linkage disequilibrium (LD) between markers was assessed by calculating squared correlation coefficients between marker intensity patterns, using 'POWERMARKER version 3.25' statistical package. The significance of pair wise LD (P values) among allele possible pairs of 55 polymorphic SNP loci was also evaluated with rapid permutation test. The plots of LD (r^2) for pairs of loci were drawn using TASSEL software.

4.3. Results

Population Structure and Relationship

The Population structure analysis was conducted using genotypic data of 55 polymorphic SNP markers by using STRUCTURE software 2.3.3 (Pritchard *et al*. 2000), and revealed that the accessions sub-divided into six subpopulations according to the suggestion of Pritchard and Wen (2007).

Cluster one contained 171 accessions which accounted 17.06 % of the total accession membership, cluster two contained 82 accessions that accounts 8.18 % of the total accessions membership, cluster three contained 229 accessions which accounts 22.85 % of the total accessions membership, cluster four contained 279 accessions which accounts 27.84 % of the total membership, cluster five contained 85 accessions which accounts 8.48 % of the total membership and the last cluster (cluster six) contained 156 accessions which account 15.57 % of the total proportion. The released varieties are clustered in group three which indicated they may have common ancestors (Table 4.5 and Fig 4.2).

Using the genotypic data, we have also analyzed the genetic distance between clusters and within clusters (nucleotide distance) using model based cluster distance calculation approach by

employing STRUCTURE software. From the six subpopulations (clusters), cluster four has large genetic distance to each of the five subpopulations. The average distance or expected heterozygosity between individuals in the same cluster revealed that cluster four (the largest cluster) showed the smallest distance (Table 4.2). But cluster five showed larger distance between individuals which may be due to some outlier genotypes present in that small group (1.39 % membership).

The population relation differentiation (F_{ST} value) of each cluster mean ranged from 0.1783 (cluster 3) to 0.9527(cluster six). It revealed high level of population fixation (F_{ST} >0.2) in each cluster except cluster three. According to Odong *et al*. (2011), weak population differentiation showed lower F_{ST} value (F_{ST} <0.05) and high level differentiation showed higher F_{ST} value (F_{ST} >0.2).The gene diversity (expected heterozygosity) is defined as the probability that two randomly chosen alleles from the population are different. It is ranged from 0.5088 (CKaM0033) to 0.0139 (CKaM0630) with average gene diversity of 0.0858 and maximum polymorphic information content of 0.388 (CKaM0033) with low heterozygosity (average of 0.0206) in the polymorphic markers. From 55 polymorphic SNP markers, 28 showed heterozygosity with the highest heterozygosity was observed in CKaM0033 (0.9790).The result also showed that 109062 total alleles were detected. The average frequency of major allele is 0.9463 with range of 0.993 (CKaM0630) to 0.5015 (CKaM0033) and the amount of heterozygosity is ranged from 0.9790 (CKaM0033) to 0.0010 (CKaM1254) (Table 4.5). The observed common allele were 1907, rare allele were 301 with unique allele six present in few genotypes.

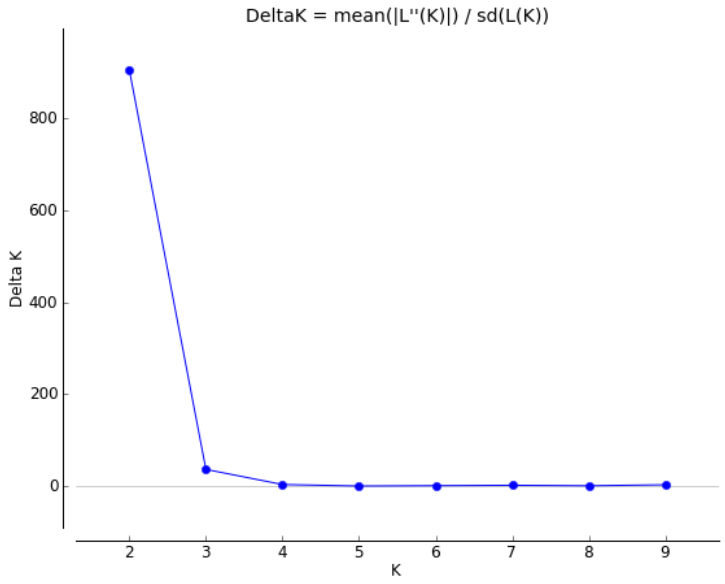

Fig.4.1 Plotting Delta K showing proper K number of clusters

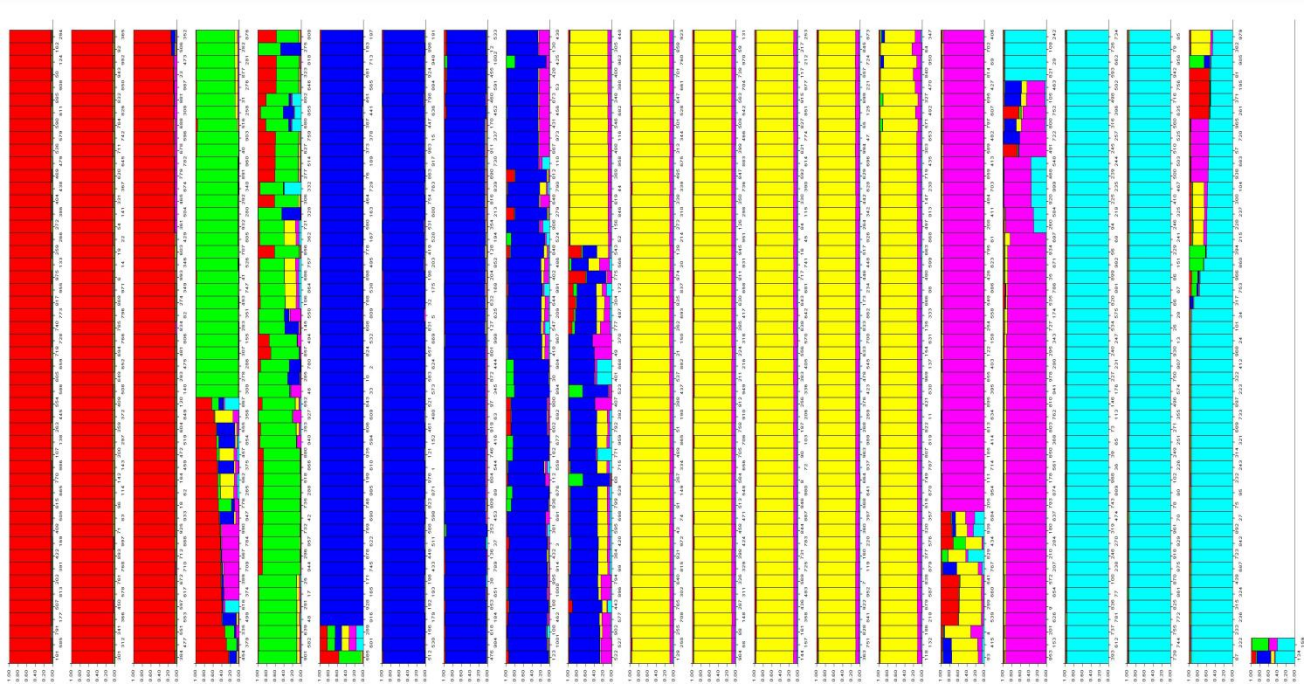

Fig 4.2 Population structure of accessions by Structure program, each individual is showing as a vertical bar partitioned in to K colored components representing inferred membership in K genetic cluster

Association analysis

According to Buckler and Thornsberry (2002), by accounting population structure with appropriate statistical methods in association test, false positive were reduced up to 80 % in a structured population.

So to reduce the risk of false positive association, we have analyzed marker- trait association using strong models (GLM and MLM). These models which are integrated in TASSEL software are using Q and K matrix to reduce false association of markers and quantitative traits by stratifying the population structure in to subpopulation using marker data for generating matrix.

The analysis of marker trait association result revealed that there was strong association between markers and agronomically important traits at two environments in the general linear model (GLM) and mixed linear model (MLM) program analysis.

In general linear model (GLM) analysis, CKaM0999 marker was highly associated with pod filling period and days to maturity at without drought stress environment in GLM and MLM analysis. At drought stressed environment, five markers were highly associated with agronomic traits such as hundred seed weight highly associated with marker CKaM0639, days to 50 % flowering highly associated with markers CKaM1140 and CKaM0888 and biomass per plot highly associated with CKaM0999 at drought stress environment in general linear and mixed linear model analysis (Table 4.3).

CKaM0999 marker showed highly significant association with pod filling period and days to maturity at non drought stress environment, and biomass per plot at drought stress environment. It showed that one marker is linked to three agronomic traits indicating that there is pleiotropy gene

controlling one or more traits or there may be allele sharing due to mutation or evolutionary the same descendent.

Linkage Disequilibrium

Linkage disequilibrium (LD) among SNP was investigated in the entire set of genotyped population and in each of the subgroups. A total of 486 pair wise computation were estimated among pairs of loci from the screened SNP markers. A total of 15 paired of loci showed statistically significant linkage disequilibrium (higher r^2 and D' values). Because allele frequency and recombination between sites affect LD, we have considered paired of loci having larger allele frequency which showed statistically significant different from zero (D value).

Since r^2 revealed both recombination and mutation history, we considered paired of loci which revealed greater r^2 value (>0.2) as in linkage disequilibrium stage (Table 4.1). The scatter plot of r^2 values also revealed linear arrangement of LD between polymorphic sites of two loci in the genomic regions (Fig.4.3).

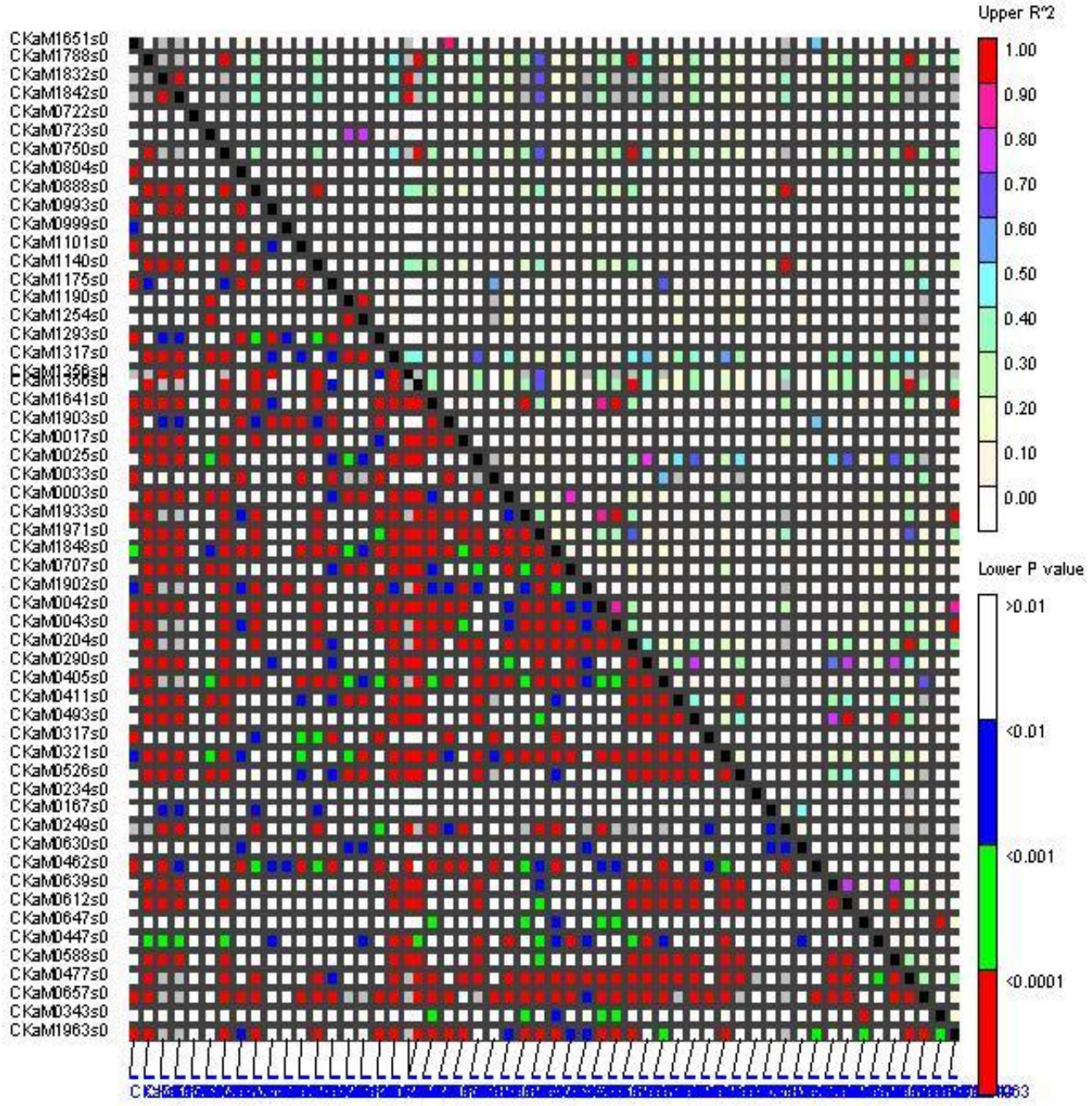

Fig 4.3 LD plot showing polymorphic sites of two loci

Table 4.1 Pairs of allele showed strong linkage disequilibrium in different loci

Sr. no.	Marker 1	Marker 2	Allele 1	Allele2	Hap frequency.	r^2	D’
1	CKaM1328	CKaM1356	G	A	0.981	0.7456	0.8767
2	CKaM1832	CKaM1842	T	C	0.9775	0.7128	0.8896
3	CKaM0639	CKaM0612	G	C	0.9815	0.5507	0.7550
4	CKaM0249	CKaM0630	C	G	0.9865	0.4074	0.8553
5	CKaM1788	CKaM1832	T	T	0.9621	0.3979	0.8879
6	CKaM0411	CKaM0493	C	T	0.9815	0.3821	0.7967
7	CKaM0321	CKaM0526	C	G	0.981	0.272	0.6791
8	CKaM1317	CKaM1328	A	G	0.9746	0.2641	0.5371
9	CKaM0447	CKaM0588	A	T	0.9820	0.2458	0.5395
10	CKaM0043	CKaM0204	C	T	0.978	0.2284	0.531
11	CKaM0588	CKaM0477	T	T	0.977	0.2274	0.6295
12	CKaM0204	CKaM0290	T	C	0.980	0.2188	0.5935
13	CKaM1933	CKaM1971	G	G	0.9795	0.2254	0.4929
14	CKaM0167	CKaM0249	T	C	0.9835	0.2170	05499
15	CKaM0477	CKaM0657	T	T	0.9581	0.2041	0.6182

Table 4.2 Cluster distance of six subpopulation and membership proportions

Cluster Group	I	II	III	IV	V	VI	Cluster group	Mean value of Fst	Membership proportion
I	**0.0133**						I	0.915	0.180
II	0.0303	**0.034**					II	0.8359	0.076
III	0.0538	0.0736	**0.1044**				III	0.1783	0.209
IV	0.0184	0.0253	0.0593	**0.0092**			IV	0.9514	0.256
V	0.0236	0.0314	0.0577	0.0071	0**.0156**		V	0.834	0.119
VI	0.0365	0.0434	0.0728	0.0182	0.0252	0**.0093**	VI	0.9527	0.160

Bold diagonal values indicate within cluster distance.

Table 4.3 List of markers and associated agronomic traits in general linear model (GLM) and mixed linear model (MLM) analysis at two environments

Sr no	Marker name	Associated traits at without drought stress environment	Traits associated to markers at drought stress environment	Model used
1	CKaM0999	PFP*** , DM***		GLM
2	CKaM0639		HSW**	GLM
3	CKaM1140		DF**	GLM
4	CKaM0888		DF**	GLM
5	CKaM0999		BM**	GLM
6	CKaM0999	PFP***, DM***		MLM
7	CKaM0888		DF**	MLM
8	CKaM1140		DF**	MLM
9	CKaM0639		HSW**	MLM

= significant (p<0.01), * = significant at (P< 0.001). DF = days to flowering, DM = days Maturity, PFP = pod filling period, HSW = hundred seeds weight, BM= biomass per plot.

Table 4.4 Distribution of accessions in each cluster, cluster based on molecular data using STRUCTURE

Cluster	Accession list
I (171)	41006, 41015, 41019, 41020, 41023, 41024, 41052, 41056, 41064, 41073, 41084, 41085, 41094, 41098, 41110, 41117, 41125, 41129, 41138, 41143, 41145, 41146, 41147, 41148, 41167, 41169, 41174, 41182, 41186, 41205, 41207, 41265, 41269, 41271, 41272, 41278, 41280, 41299, 41300, 41303, 41307, 41314, 41319, 207142, 207146, 207153, 207160, 207163, 207166, 207171, 207174, 207176, 207180, 207181, 207182, 207188, 207192, 207195, 207564, 207611, 207613, 207616, 207618, 207631, 207635, 207636, 207650, 207657, 207667, 207674, 207680, 207684, 207687, 207689, 207695, 207700, 207703, 207704, 207706, 207709, 207711, 207713, 207732, 207739, 207741, 207742, 207743, 207744, 207759, 208829, 208991, 208993, 208994, 208997, 209004, 209010, 209021, 209023, 209030, 209031, 209036, 209082, 209083, 209089, 209090, 209091, 209099, 209104, 209105, 209106, 209109, 214621, 214626, 214730, 215032, 215289, 216855, 219797, 219801, 223143, 225872, 225888, 225889, 225890, 227149, 227155, 228198, 228296, 228298, 229960, 230253, 231327, 231329, 231331, 232206, 232287, 233353, 233572, 233750, 234048, 235033, 235393, 235826, 236454, 236461, 236471, 236477, 236479, 236887, 237054, 237513, 239846, 239849, 239876, 239879, 239890, 239906, 239927, 239938, 239959, 239966, 239979, 240063, 240067, 240078, 240089, 240090, 241127, 241800, 241801, 241805
II (82)	41018, 41026, 41032, 41041, 41042, 41043, 41047, 41049, 41109, 41150, 41160, 41264, 41281, 41282, 41283, 41284, 41286,41287, 41288, 41289, 41291, 41292, 41294, 41295, 41297, 41298, 41302, 41312, 41313, 41315, 207133, 207139, 207143, 207152, 207165, 207167, 207177, 207726, 207727, 207750, 207752, 207754, 207767, 208990, 208992, 209027, 209087, 209116, 212689, 214622, 219803, 223064, 223065, 225886, 227160, 228288, 228291, 228303, 229955, 229958, 233570, 235720, 235825, 236193, 236462, 238262, 238264, 238265, 239842, 239845, 239859, 239860, 239863, 239888, 239898, 239901, 239929, 239937, 239950, 239960, 239982, 240043, 41294, 41295, 41297, 41298, 41302, 41312, 41313, 41315, 207133, 207139, 207143, 207152, 207165, 207167, 207177, 207726, 207727, 207750, 207752, 207754, 207767, 208990, 208992, 209027, 209087, 209116, 212689, 214622, 219803, 223064, 223065, 225886, 227160, 228288, 228291, 228303, 229955, 229958, 233570, 235720, 235825, 236193, 236462, 238262, 238264, 238265, 239842, 239845, 239859, 239860, 239863, 239888, 239898, 239901, 239929, 239937, 239950, 239960, 239982, 240043
III (229)	41001, 41002, 41003, 41005, 41010, 41012, 41016, 41031, 41033, 41034, 41038, 41039, 41051, 41055,41062, 41065, 41078, 41091, 41099, 41102, 41111, 41113, 41115, 41126, 41128, 41132, 41133, 41135, 41141, 41157, 41168, 41170, 41171, 41172, 41173, 41175, 41176, 41177, 41180, 41184,41187, 41188, 41189, 41193, 41194, 41196, 41197, 41198, 41199, 41201, 41202, 41203, 41204,41208, 41209, 41214, 41218, 41258, 41285, 41310, 207149, 207159, 207168, 207179, 207191, 207563, 207607, 207614, 207619, 207628, 207629, 207634, 207644, 207646, 207647, 207648, 207653, 207656, 207658, 207659, 207661, 207662, 207670, 207672, 207673, 207676, 207679, 207681, 207682, 207683, 207685, 207686, 207690, 207691, 207692, 207694, 207707, 207712, 207718, 207719, 207728, 207747, 207748, 207760, 207761, 207763, 207764, 207765, 207768, 207770, 207894, 207895, 207896, 208364, 208977, 208982, 208983, 208986, 209001, 209008, 209009, 209016, 209020, 209025, 209026, 209028, 209033, 209038, 209080, 209081, 209084, 209086, 209088, 209092, 209094, 209096, 209098, 209107, 209111, 209113, 209114, 212476, 212686, 213224, 214625, 214729, 214916, 215353, 215577, 215667, 216856, 219799, 219800, 223288, 225740, 225742, 225876, 227150, 227152, 228199, 228200, 228301, 228302, 228658, 229090, 229962, 231330, 231332, 232286, 232288, 233352, 234852, 235031, 235032, 235034, 235036, 235395, 235396, 235398, 235722, 236194, 236198, 236459, 236468, 236469, 236476, 236488, 236493, 236886, 237056, 237057, 239841, 239851, 239852, 239861, 239874, 239886, 239891, 239893, 239895, 239897, 239900, 239902, 239904, 239907, 239909, 239914, 239916, 239917, 239919, 239925, 239930, 239941, 239945, 239967, 240042, 240058, 240065, 240068, 240085, 240087, 240088, 241128, 241803, 241804, 207610re, Fetenech, kutaye, Mariye
IV (279)	41004, 41007, 41008, 41011, 41017, 41021, 41022, 41044, 41045, 41046, 41048, 41053, 41054, 41057, 41060, 41061, 41066, 41074, 41076, 41086, 41090, 41092, 41093, 41095, 41101, 41106, 41119, 41120, 41121, 41123, 41124, 41130, 41131, 41134, 41136, 41137, 41140, 41142, 41144, 41149, 41152, 41153, 41154, 41155, 41159, 41161, 41162, 41165, 41166, 41178, 41191, 41192, 41195, 41211, 41216, 41217, 41219, 41221, 41222, 41224, 41225, 41226, 41229, 41236, 41240, 41256, 41259, 41261, 41262, 41268, 41270, 41273, 41274, 41275, 41279, 41293, 41304, 41305, 41308, 41309, 41311, 41316, 41317, 41320, 41325, 202509, 207136, 207138, 207140, 207141, 207144, 207145, 207147, 207150, 207151, 207155, 207157, 207161, 207162, 207164, 207172, 207173, 207175, 207178, 207184, 207561, 207562, 207606, 207609, 207612, 207615, 207617, 207624, 207625, 207626, 207627, 207632, 207643, 207645, 207649, 207651, 207652, 207663, 207664, 207665, 207666, 207669, 207671, 207675, 207677, 207688, 207696, 207701, 207702, 207710, 207715, 207717, 207720, 207721, 207723, 207725, 207730, 207734, 207745, 207749, 207751, 207753, 207769, 208900, 208978, 208980, 208981, 208984, 208985, 208988, 208998, 208999, 209006, 209007, 209011, 209012, 209013, 209014, 209019, 209032, 209076, 209085, 209103, 209108, 209110, 210858, 210859, 212474, 212475, 212477, 212688, 212916, 212917, 213050, 213051, 214624, 214731, 214734, 214917, 215033, 215067, 215189, 215190, 215290, 216853, 219798, 219802, 219804, 221696, 223063, 225743, 225874, 225877, 225878, 225879, 225882, 225883, 225884, 225887, 227154, 227156, 227970, 227971, 228201, 228292, 228294, 228297, 228792, 228899, 229956, 229959, 230795, 231328, 232207, 233573, 234050, 235963, 236197, 236458, 236463, 236467, 236472, 236473, 236474, 236475, 236478, 236480, 236482, 236483, 236489, 236491, 236881, 236882, 236883, 236884, 237055, 238267, 239840, 239844, 239847, 239853, 239862, 239865, 239873, 239878, 239908, 239911, 239912, 239915, 239922, 239923, 239924, 239928, 239936, 239946, 239949, 239961, 239965, 239968, 239971, 239972, 239976, 240041, 240045, 240048, 240049, 240060, 240061, 240064, 240066, 240069, 240077, 241126, 241802,

V (85)	41009, 41036, 41063, 41071, 41108, 41114, 41127, 41158, 41163, 41179, 41183, 41185, 41190, 41206, 41210, 41212, 41213, 41215, 41260, 41266, 41290, 41296, 41301, 207156, 207183, 207185, 207622, 207630, 207633, 207638, 207641, 207642, 207654, 207655, 207693, 207699, 207714, 207716, 207724, 208454, 208979, 208989, 209000, 209003, 209015, 209029, 209034, 209093, 209102, 209112, 209115, 212589, 212687, 214728, 214732, 214733, 216854, 225873, 225875, 225880, 225881, 227151, 227161, 228197, 228900, 229961, 234049, 235035, 235392, 235394, 235397, 235721, 236196, 236457, 239253, 239855, 239877, 239892, 239921, 239939, 239954, 239977, 239978, 240071, 240062re
VI (156)	41014, 41025, 41027, 41028, 41029, 41030, 41035, 41037, 41040, 41058, 41059, 41067, 41068, 41069, 41070, 41072, 41075, 41077, 41079, 41080, 41081, 41082, 41083, 41087, 41088, 41089, 41096, 41097, 41103, 41104, 41105, 41107, 41112, 41116, 41139, 41151, 41156, 41164, 41181, 41200, 41220, 41223, 41227, 41228, 41230, 41231, 41232, 41233, 41235, 41237, 41238, 41239, 41241, 41242, 41243, 41244, 41245, 41246, 41247, 41248, 41249, 41250, 41251, 41252, 41253, 41254, 41255, 41257, 41263, 41267, 41276, 41277, 41306, 41321, 41322, 41323, 41324, 41326, 204785, 205148, 207134, 207135, 207170, 207186, 207608, 207620, 207621, 207623, 207637, 207639, 207668, 207698, 207705, 207729, 207731, 207733, 207735, 207736, 207746, 207766, 208453, 209002, 209017, 209018, 209078, 209101, 212685, 215188, 222863, 223142, 223287, 227153, 227157, 227162, 228196, 228289, 228290, 228293, 228295, 228299, 228300, 228793, 229089, 229091, 229703, 231454, 233571, 236464, 236465, 236470, 236481, 236492, 239836, 239838, 239850, 239857, 239864, 239885, 239889, 239894, 239896, 239903, 239905, 239918, 239932, 239935, 239947, 239957, 239980, 240044, 240054, 240055, 240059, 240073, 240080, 240091

Number in parentheses indicated number of accessions in each cluster

Table 4.5 Summary of allele frequency, gene diversity, heterozygosity and polymorphic information content (PIC) value for polymorphic markers.

Sr no	Marker name	Major allele frequency	Gene diversity	Heterozygosity	PIC value
1	CKaM0003	0.9880	0.0237	0.0000	0.0236
2	CKaM0017	0.9731	0.0527	0.0040	0.0519
3	CKaM0025	0.9815	0.0364	0.0010	0.0360
4	CKaM0033	0.5015	0.5088	0.9790	0.3882
5	CKaM0042	0.9815	0.0364	0.0110	0.0360
6	CKaM0043	0.9870	0.0257	0.0000	0.0255
7	CKaM0167	0.9910	0.0178	0.0000	0.0177
8	CKaM0204	0.9840	0.0315	0.0000	0.0313
9	CKaM0234	0.9870	0.0256	0.0000	0.0254
10	CKaM0249	0.9875	0.0247	0.0010	0.0245
11	CKaM0290	0.9900	0.0198	0.0000	0.0197
12	CKaM0317	0.7974	0.3273	0.0120	0.2805
13	CKaM0321	0.9840	0.0316	0.0020	0.0313
14	CKaM0343	0.9800	0.0392	0.0000	0.0387
15	CKaM0405	0.9691	0.0604	0.0000	0.0594
16	CKaM0411	0.9835	0.0325	0.0010	0.0322
17	CKaM0447	0.9870	0.0257	0.0000	0.0255
18	CKaM0462	0.8293	0.2886	0.0140	0.2560
19	CKaM0477	0.9810	0.0374	0.0000	0.0370
20	CKaM0493	0.9900	0.0198	0.0000	0.0197
21	CKaM0526	0.9905	0.0188	0.0010	0.0187
22	CKaM0588	0.9890	0.0218	0.0000	0.0216
23	CKaM0612	0.9850	0.0296	0.0000	0.0293
24	CKaM0630	0.993	0.0139	0.0000	0.0138
25	CKaM0639	0.9855	0.0286	0.0010	0.0284
26	CKaM0647	0.9910	0.0178	0.0000	0.0178
27	CKaM0657	0.9651	0.0680	0.0000	0.0668
28	CKaM0707	0.9875	0.0247	0.0010	0.0246
29	CKaM0722	0.992	0.0159	0.0000	0.0158
30	CKaM0723	0.9840	0.0315	0.0020	0.0312
31	CKaM0750	0.9760	0.0470	0.0000	0.0465
32	CKaM0804	0.9022	0.1784	0.0100	0.1659
33	CKaM0888	0.9850	0.0296	0.0000	0.0293
34	CKaM0993	0.9182	0.1526	0.0060	0.1451
35	CKaM0999	0.9431	0.1083	0.0020	0.1043
36	CKaM1101	0.8598	0.2435	0.0110	0.2179
37	CKaM1140	0.9750	0.0489	0.0000	0.0480
38	CKaM1175	0.9800	0.0393	0.0000	0.0388
39	CKaM1190	0.9825	0.0344	0.0010	0.0340
40	CKaM1254	0.9895	0.0208	0.0010	0.0207
41	CKaM1293	0.6467	0.4709	0.0140	0.3780
42	CKaM1317	0.9820	0.0354	0.0000	0.0351
43	CKaM1328	0.9835	0.0325	0.0010	0.0322
44	CKaM1356	0.9830	0.0335	0.0000	0.0332
45	CKaM1641	0.9721	0.0547	0.0000	0.0539
46	CKaM1651	0.7520	0.3833	0.0170	0.3254
47	CKaM1788	0.9641	0.0698	0.0000	0.0684

48	CKaM1832	0.9815	0.0364	0.0010	0.0361
49	CKaM1842	0.9795	0.0403	0.0010	0.0399
50	CKaM1848	0.9541	0.0885	0.0000	0.0862
51	CKaM1902	0.9077	0.1695	0.0050	0.1586
52	CKaM1903	0.7635	0.3678	0.0200	0.3103
53	CKaM1933	0.9870	0.0257	0.0000	0.0255
54	CKaM1963	0.9780	0.0432	0.0000	0.0427
55	CKaM1971	0.9860	0.0276	0.0120	0.0275
	Mean	0.9463	0.0858	0.0206	0.0769

4.4 **Discussion**

Association mapping (AM), based on linkage disequilibrium, is a complementary strategy to traditional quantitative trait loci (QTL) mapping for describing associations between genotypes and phenotypes in crop plants. It was suggested as a promising alternative strategy to linkage mapping elucidating the genetic basis of complex traits. Population stratification, information on the relatedness among genotypes is commonly incorporated as population effect or as kinship matrix to correct nonfunctional association between the traits under consideration and the underlying population structure (Yu *et al*. 2006). One popular method correcting population stratification is using Bayesian model based framework implemented in the software package STRUCTURE (Yu *et al.* 2006).

This finding is expected to show distinct subpopulation structure with strong population differentiation and fixation (high mean Fst value) even if it could not show distinct heterotic group like hybrid breeding populations. The population structure analysis revealed higher mean Fst value indicating that the stratified subpopulations were strongly fixed and significantly different from the original population. Highest Fst value observed in cluster IV and cluster VI which contributed 25.6 % and 16 % of total membership, respectively. We also observed better genetic distance between subpopulations especially between cluster III and VI. Since the distribution of functional alleles is highly correlated with population structure, better subpopulation differentiation reduced the rate of

false positive association between markers and studied traits. Inclusion of population subdivision as random effects in a mixed model allows for the computation of unbiased estimates of allele effect.

We observed high number of SNP marker pairs in linkage disequilibrium revealed that there is strong correlation between alleles either due to physical linkage on the same chromosome or due to sharing of alleles which were identical by decent or due to mutation and evolution process.

The association analysis result revealed that markers are strongly associated with days to 50 % flowering and days to maturity at both GLM and MLM analysis. This indicated that these markers are derived from regions coding drought resistance. CKaM0999 marker strongly associated with pod filling period, days to maturity and biomass in both MLM and GLM analysis at both droughts stressed and without drought stress environment with major allele frequency of 0.9431.

CKaM0033 with better PIC value (0.38) is associated with pod per plant and seeds number per plant at without drought stress environment in GLM and MLM analysis. Since the study germplasm is diverse, association analysis captured more allelic diversity with better resolution without any nonfunctional association.

Once the genetic markers have been demonstrated to be associated with a phenotypic trait of interest, it can be used as selection target to obtain an indirect response in the trait. In recurrent selection, markers could be used to store information acquired from phenotypic evaluations, which can be used for selection in later cycles. Likewise, in pedigree breeding, markers could carry information about yield potential from the phase of replicated field trials to the phase of single-plant selection, when evaluation of yield cannot be made with reasonable precision. It is also useful for the breeders to select exclusively the favorable marker allele, trying to achieve fixation

of the favorable gene allele in a single generation. So this finding is useful for indirect selection for traits strongly associated with markers. Similar findings reported in maize (Setter *et al*. 2011).

5. Developing Diversity Based Chickpea (*Cicer arientinum* L.) Core Collection to Foster Germplasm Utilization

5.1. Introduction

Plant genetic resource refers to the sum total of genes, gene combinations or genotypes available for the genetic improvement of crop plants. Plant genetic resources will be the main contributing factor to future progress in developing new cultivars (Upadhyaya *et al*. 2007). Germplasm collections were originally set up to preserve the genetic diversity of crop species and their wild relatives.

Given that such genetic diversity of crops has an economic value, conservation for use has been the driving force behind many gene banks. But the sheer number of accessions and the resultant costs of their maintenance making up germplasm collections could be an obstacle for their full exploitation, evaluation and utilization to impact the crop improvement or breeding programmes (Franco *et al*. 2006). In this regard, genetic diversity of such a large collection may not have been adequately evaluated for various biotic and abiotic stresses, due to resource and time constraints. It is impractical to evaluate such large collections in detail as it would be expensive and time consuming. Selecting a few lines from these vast pools of germplasm is like searching for a needle in a haystack. Obviously, it is more appropriate and attractive to have a small sample of a few hundred germplasm lines, based on critical evaluation, representing the entire diversity of the species. So genetic resources stored in gene banks are usually sampled to foster efficient evaluation and utilization of the collections as well as to study phenotypic and genotypic diversity, from subsets, and eliminate redundant and duplicate accessions.

This task could be more easily fulfilled by developing subsets of the whole collection, called active working collections by Harlan (1972) and core collections by Frankel and Brown (1984).

The main purpose of developing core collection is to preserve in the sample as much of the diversity present in the original collection as possible (Crossa *et al*. 1995a). For example, the approach of forming core collections (core subsets) was introduced to increase the efficiency of describing and using collections stored in gene banks, while preserving as much as possible the diversity of the entire collection (Frankel and Brown, 1984; Brown, 1989).

A core collection (called also a "core subset") derived from an existing entire collection (a gene bank) within a crop species should include a maximum of the genetic variation available in the whole collection with minimal repetitiveness, ideally conserving at least 70% of the alleles in the whole collection (Brown, 1989a, 1989b). Then, the core collection consists of a limited number of the accessions from the existing collection that represent the diversity (or spectrum) in the entire collection. Representativeness is the most important property for a core collection. It is defined as similarity of the genotypic or/and phenotypic diversity in a core collection with the respective diversity in the entire collection.

Several statistical methods, referred to as sampling strategies or sampling methods, have been introduced for the selection of accessions from an existing genetic resources collection to form a core collections that are as representative as possible (Upadhyaya *et al*. 2007; Wang *et al*. 2007). These methods include simple random sampling (Brown, 1989) and stratified random sampling (Franco *et al*. 2006; Xu *et al*. 2006; Wang *et al*. 2007).

The process of stratified random sampling of genetic resources with the objective of forming subsets starts with determining the size of the core, stratify or grouping accessions to obtain

homogenous within and heterogeneous between clusters (groups) and then using a predetermined sampling strategy select at random entries within each cluster to form core set.

The grouping of accessions in to clusters is achieved by a classification strategy that partitions the original collections in to groups with maximum distances between accessions located in different groups and minimum distances between accessions located in the same group Franco *et al*. (1998, 1999, 2002). The sampling intensity of core collection development ranged from 5% - 20 % of the total number of accessions. This intensity of sampling captures 86-90% of the diversity present in the reserve collections (Brown, 1989).

Core collections for national or global diversity have been established for many crop species, using morphological and genetic marker variation; e.g. bean Phaseolus *vulgaris* L. (Tohm *et al*. 1995); barley Hordeum *vulgare* (Knu¨pffer and van Hintum, 1995); chickpea Cicer *arietinum* L. (Hannan *et al*. 1994; Upadhyaya *et al*. 2001).

So the objective of this research was to develop core collection that represent the genetic diversity of entire collection of Ethiopian chickpea germplasm preserved in the Institute of Biodiversity and Conservation gene bank with a minimum reputation, to facilitate germplasm utilization in the national chickpea improvement program and to identify accessions tolerant to drought stress and other desirable agronomic traits in the core set.

5.2. Materials and Method

Plant Material and experiment layout

The Ethiopian chickpea germplasm collections which were collected by Institute of Biodiversity and Conservation (IBC) Ethiopia from the whole country since 1970 were considered for study. The entire collection contains 1157 accessions with 99 % desi type and very little kabuli type (1 %).

From these entire collections, 1035 chickpea accessions, which are desi type and produced enough seed in seed multiplication, were evaluated for eleven quantitative traits at Kobo research site (representing drought prone area in north- eastern Ethiopia). Genotypes were planted in two environments (drought stress and without drought stress environments) in 2010 main cropping season. Even though chickpea is considered a drought tolerant crop, its seed yield can increase also with a supplementary irrigation, applied between flowering and seed growth (Soltani *et al*. 2001). So without drought stress environment was created by applying supplemental irrigation at flowering and pod setting time of the crop. The experiment was laid in randomized complete block design with two replications at both environments. Data were recorded for the eleven quantitative traits.

So at the moment of this work, 1035 accessions' quantitative data were used for validating and developing core set of Ethiopian chickpea germplasm collections.

From 1035 agronomically evaluated accessions, 1002 accessions were also genotyped for generating SNP data. 111 SNP markers that covered the whole linkage group of chickpea were used for genotyping 1002 genotypes. From the screened SNP markers, 55 SNP markers revealed polymorphism in one or more genotypes. These SNP data were used for diversity analysis and developing core collection that represent the diversity of chickpea collection present in the Institute of Biodiversity Conservation gene bank of Ethiopia.

Sampling Strategies for Constructing the Core Collection

The procedure used to establish the Ethiopian chickpea core collection was based on the concept that preexisting informations (passport data) about the collection were used to stratify the accessions. In this work, the sampling procedure followed the general procedure suggested by

Brown (1989a). So the whole germplasm collection present in the Institute of Biodiversity and Conservation gene bank of Ethiopia was stratified by agro-ecology or geographic distributions. Based on these stratifications, genotypes grouped in to five major clusters considering that accessions from wide geographical area will provide indirect evidence of diversity since accessions from the same origin (state) can be assumed to share a large portion of their gene pool (Peeters and Martienlli, 1989). Based on genotypic data (SNP data), cluster analysis was done using DARwin 5.0 software for each five group. Samples were drawn randomly from each cluster group in each major growing areas based on number of genotypes (proportion) and diversification.

The sampling of entire collection included the following steps

1. It was decided that 15 % of the whole would represent the working collection (core collection)
2. A proportional method adjusted by the relative importance of the chickpea growing area (state) was used to select the accession
3. Representative samples were taken from the stratified accession based in each geographical origin.
4. Random selection of representative accessions was used from each cluster in each growing areas.

The number of accessions selected from each state or geographical area to form core set collection was proportional to the size of the group and their geographical distribution in the whole collection. So this approach ensures that each group is represented in the core collection according to its proportion in the whole collection.

Chickpea growing area (geographical condition) in the country includes Amara region contained 468 accessions, Oromiya region contained 294 accessions, Southern Nations and Nationalities Peoples region (SNNP) contained 45 accessions, Tigray region including Eritra collection contained 68 accessions, Unknown (no information about collection region) contained 154 accessions, Somalia region contained 3 accessions, Gambela region contained 1 accession and Benishangul Gumz region contained 2 accessions. The number of accessions collected in these regions differs based on the diversity of agro-ecology (soil type and altitude), size of the region and climatic conditions in that region which favored more diversity through time. The last three regions showed very few accessions which were difficult to select based on random sampling.

The Indices for Evaluating Representativeness of Core Collection

For the developed core collections, two indices of validities (goodness or quality in a sense of representativeness) were used (Kim *et al*. 2007). The first index refers to the average of absolute differences between means across all of the traits in the core and entire collections relative to the means in the entire collection, MD %. The other index is the average of the absolute differences between variances across all of the traits in the core and entire collections relative to the variance in the entire collection, VD %. The goodness indices were calculated according to the formulas (Kim et al. 2007):

$$MD\ \% = \frac{\sum_{t=1}^{p} \left| \bar{x}_{Ct} - \bar{x}_{Et} \right|}{\frac{\bar{x}_{Ct}}{P}} \times 100$$

$$VD\ \% = \frac{\sum_{t=1}^{p} \left| \sigma^2_{Ct} - \sigma^2_{Et} \right|}{\frac{\sigma^2_{Et}}{P}}$$

Where $\bar{x}_{Ct}$ = the mean of the i^{th} (t = 1, 2, ---p) trait for a core collection, $\bar{x}_{Et}$ = the mean of the i^{th} trait for the entire collection, σ^2_{Ct} = the variance of the i^{th} trait for a core collection σ^2_{Et} = the variance i^{th} trait for entire collection

Smaller values of MD % and VD % for the sampling strategy indicate a more effective strategy, e.g. smaller values show a better ability of the sampling strategy to establish a representative core collection. The coincidence rate of percentage (CR %) and percentage variable rate of coefficient of variance (VR%) were calculated based on the formula Hu *et al*. (2000) to measure the percentage of the significant difference of traits between core and entire collections:

CR % = (1/m) $\sum$ (R_C /R_1) X 100, where R_C = Range of the core collection, R_1= Range of the initial collection and m = number of traits.

VR % = (1/m) $\sum$ (CV_C /CV_1) X 100 where CV_C Coefficient of variation of the core collection, CV_1 = the coefficient of variation of the initial collection, m = number of traits

Identification of Desirable Genotypes from Core Set

The entire genotypes were evaluated at drought stressed and non drought stressed environments and observation was done on eleven agronomic traits such as days to 50 % flowering, days to maturity, plant height, biomass per plot (kg), hundred seeds weight, seed number per plant, grain yield per plot (gm), grain yield per plant, pod filling period, harvest index and pod per plant and validation of core set was done using these agronomic traits.

Tolerance to drought stress (TDS), mean productivity (MP), drought tolerance efficiency (DTE) and rate of productivity (RP) of selected genotypes for core set were calculated based on the formula suggested by Rosiell and Hambin (1981). Where TDS = Y_2-Y_1. TDS = tolerance to drought stress, Y_1 = seed yield in the non drought stress environment and Y_2 = seed yield in the

stress environment. Mean productivity was calculated by using the formula: MP = $(Y_1 + Y_2)/2$. The mean productivity was defined as MP and rate of productivity arbitrated as RP. Where RP = (Y_2/Y_1) and DTE = RP *100. Drought susceptibility index (DSI) was also calculated with the formula suggested by Fischer and Maurer (1978). DSI = 1- (Y_2/Y_1) /D where D = the ratio of mean of all the genotypes in Y_2 and mean of all the genotypes in Y_1. Correlation analysis of core set was done using Agrobase V.33 to compare the trends of correlated traits in the entire set and core set.

Statistical analysis and computation

The statistical analyses of agronomic data were carried out using SAS V 9.1 (SAS Institute 2004) Agrobase V. 33 and Darwin 5.0 statistical packages. Hierarchical clustering using Ward method was used to depict distinct clusters from each geographical region. The genotypic data analysis was done using the software Structure V. 2.3.3, TASSEL V. 2.1 and Powermarker V. 3.25.

5.3. Results

Core set Development

According to the standard procedure indicated by Brown (1989), the entire collections were stratified in to five major groups and three small groups based on passport data. The major groups were further hierarchically clustered based on Ward method using SNP genotypic data. The hierarchical cluster analysis revealed distinct cluster group for each major groups (Table 5.1).

Using proportional sampling method and 15 % sampling intensity, 154 accessions from major groups in each cluster and 4 accessions from minor group were selected randomly without replacement within cluster members for core collection development. Since the number of

accessions in the minor group was very small, the accession selected for core set development were not proportional to 15 % selection intensity (Table 5.1). The hierarchical clustering assisted for grouping similar accessions together with in each geographical region and from each cluster group at least one accession was chosen to ensure all the cluster groups were included in the core set proportionally.

The analysis of Shannon-Weaver diversity index for 11 quantitative traits data revealed high genetic diversity between accessions in the entire collection. This genetic diversity of entire collection was observed in the core set collection. The Shannon-Weaver diversity index (H) was more than one for all the evaluated traits in both the core set and entire set (Table 5.17).

Validating the core collection sampling

The analysis of variance for core set and entire set revealed that there was highly significant difference between genotypes in biomass per plot and harvest index at without drought stress environment in the core set and entire set collection. Significant difference was also observed in days to maturity and pod filling period. Non-significant difference also observed in the traits like days to 50 % flowering, plant height and hundred seeds weight for core set and entire collection. In drought stress environment, there was highly significant difference between genotypes for yield per plot and biomass per plot. The same trend was observed between core set and entire set in most of the traits except days to 50 % flowering, seeds number per plant and plant height (Table 5.1 and 3.1).

Table 5.1 Summary of analysis of variance for eleven traits at two environments

S.O.V	**Mean of Square at without drought stress environment**										
	BM	DF	HI	DM	PFP	PHT	PPP	SPP	HSW	YLD	YPP
Block	0.29ns	23.405ns	0.009ns	0.155ns	27.37ns	171.801**	16.405ns	225.598ns	1.025ns	200.01ns	15.85ns
Genotype	0.211**	6.748ns	0.006**	8.911*	8.82*	19.371ns	137.412**	0.226.838*	2.169ns	5898.849*	5.952ns
Residual	0.218	7.042	0.005	6.353	9.186	22.138	134.118	201.012	1.885	4100.034	5.449
CV %	31.81	5.08	30.64	2.65	7.09	10.02	20.69	18.71	9.13	19.73	17.32
S.O.V	**Mean of Square at drought stress environment**										
	BM	DF	HI	DM	PFP	PHT	PPP	SPP	HSW	YLD	YPP
Block	0.297*	626.661**	0.011*	10.282ns	797.481**	0.456*	1803.877**	28.56ns	14.206**	2246.222ns	25.781*
Genotype	0.135**	9.443*	0.003*	5.226ns	7.767ns	30.27ns	448.131ns	191.226*	1.873ns	5164.525 **	7.368*
Residual	0.065	9.037	0.002	4.307	8.819	26.602	229.934	174.025	1.48	2501.418	5.113
CV %	19.06	6.1	22.35	2.23	6.81	11.26	23.39	17.25	8.03	18.2	15.32

S.O. V = Source of variance, BM = biomass per plot, DF = days to 50 % flowering, HI = harvest index, HSW = hundred seeds weight, DM = days to maturity, PFP = pod filling period, PHT = Plant height, PPP = pod per plant, SPP = seeds number per plant, YLD = yield per plot, YPP = yield per plant, * = significant at ($p< 0.05$) and ** = significant at ($p< 0.01$)

The percentage mean difference (MD %), percentage variance difference (VD %), the coincidence rate of range (CR %) and variable rate of coefficient of variance (VR %) were not significantly different (difference < 20 % for MD and VD) between core set and entire set in both environments (Table 5.8). The variability coincidence rate and variable rate of coefficient of variance is greater than 87 % and 100 %, respectively for most of the traits.

Result of correlation analysis of core set revealed that there was highly positive correlation between yield per plot and other evaluated traits except pod filling period. Biomass per plot also revealed highly significant positive correlation with the evaluated traits except pod filling period and yield per plant. Harvest index was also positively correlated only with biomass per plot, yield per plot and hundred seeds weight at without drought stress environment (Table 5.13). At drought stressed environment, yield per plot and biomass per plot showed highly positive correlation with most of the studied traits except days to maturity and plant height. Harvest index

also showed significant correlation with other traits except maturity days and pod per plant (Table 5.14). So the traits correlation analysis at core set and entire collections was similar indicating that the diversity present in the entire set is correctly represented in the core set. So the core set validation parameter revealed that the established core set represent the genetic diversity of Ethiopian chickpea entire collection properly.

Desirable Genotypes Identification

Since terminal drought stress is one of the production limiting factors for chickpea yield in Ethiopia, identifying desirable genotypes from the core set was the most important activity in this finding. So one of the criteria to identify desirable genotypes for drought stress resistance was yield stability across the stress and non drought stress environment. Yield stability, or the extent of variation in yield between stress and non stress conditions, is widely accepted as an indicator of genotypic response to stress (Blum, 1988).

The genotypes in the core set revealed yield stability for the drought stress and non stress environments. Around 30 accessions showed better performance and revealed good result in tolerant to drought stress (TDS) compared to the other accessions from the core set at drought stress environment. They also revealed good drought tolerance efficiency (DTE) compared to other genotypes. These accessions also showed lowest drought susceptibility index (DSI) which is very important parameter for screening drought tolerant genotypes. Tolerant to drought stress ranged from -66.80 (Fetenech) to 182.25 (41034), drought susceptibility index ranged from -1.019 (41034) to 0.237 (Fetenech) and drought tolerance efficiency ranged from 79.883 (Fetenech) to 186.58 (41034).

Some researchers announced that the cultivars which had the lowest DSI values were drought resistant than the cultivars with the highest DSI values (Zerea-Fizabady and Ghodsi, 2004; Golabadi et al. 2006). So the minimum yield reduction was realized in the accessions which had the highest DTE and the lowest DSI (Table 5.7). Compared to the best performing check (Kutaye) in the drought stress environment, 18 accessions showed better drought tolerance efficiency with lower drought susceptibility index.

Genetic Diversity of Core Set

The phenotypic data based hierarchical cluster analysis of 158 core set showed five distinct clusters in the non drought stress environment and six distinct groups at drought stress environment. The number of accessions in each distinct group ranged from 2 accessions in the 1^{st} group to 63 accessions in the 5^{th} group at without drought stressed environment. At drought stress environment, the number of accessions in each cluster group ranged from 2 accessions at 3^{rd} and 5^{th} clusters to 59 accessions at 4^{th} cluster (Fig.5.1& 5.2)

Fig.5.1 Schematic representation of phenogram at without drought stress environment, blue = cluster 1, red cluster 2, rose = cluster 3, light blue = cluster 4 and black cluster 5

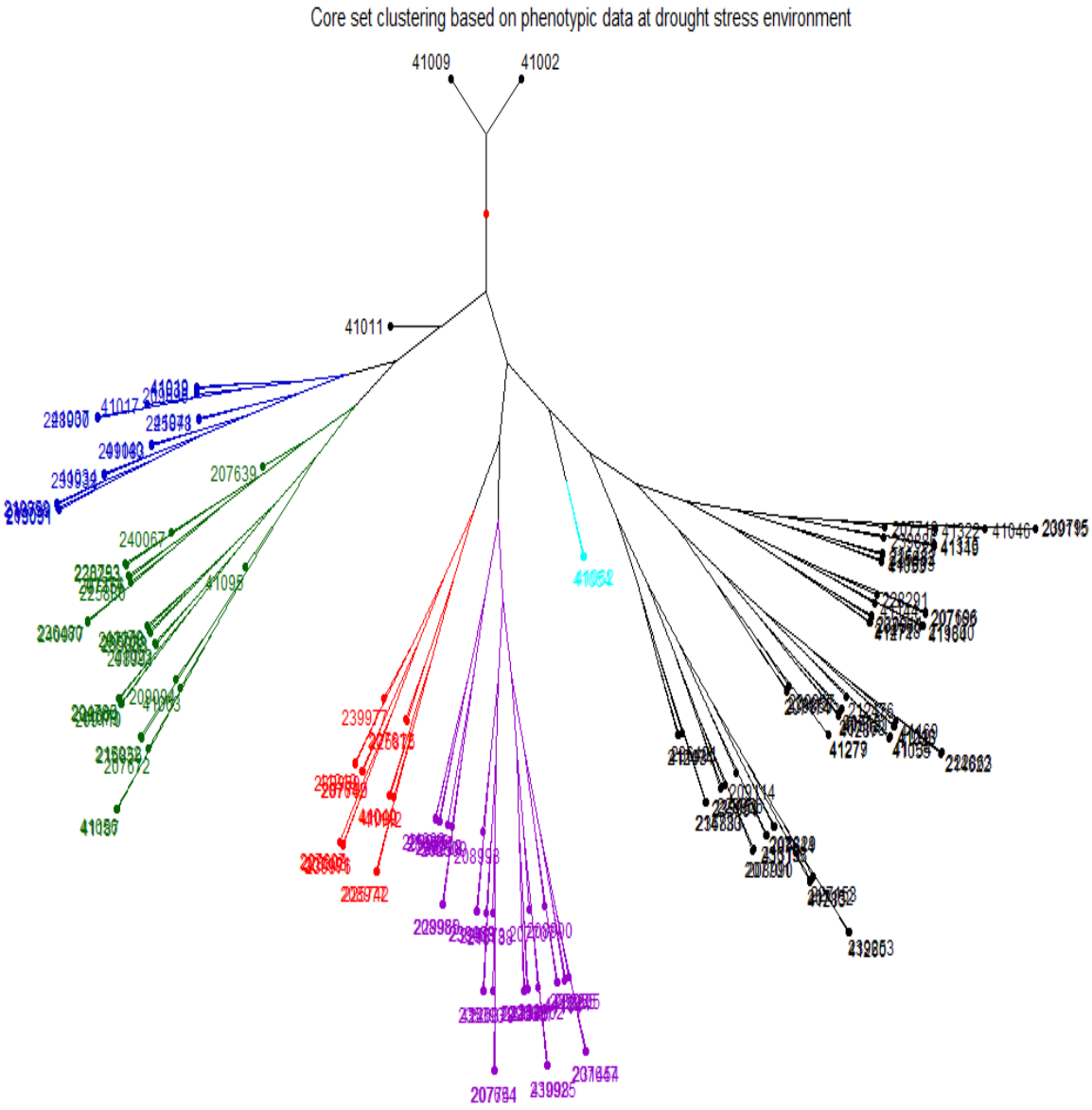

Fig. 5.2 Schematic representation of phenogram at without drought stress environment, blue = cluster 1, green = cluster 2, red = 3, rose = cluster 4, light blue = cluster 5 and black = cluster 6; 41009, 41002 and 41011 out of any cluster

Analysis of cluster distance between groups revealed better diversity in the core set. The distance between clusters ranged from 3.40 (between cluster 2 and 5) to 21.80 (between cluster 1 and 5) which showed cluster 1 and 5 are more distant genetically than other group distance at without drought stress environment. At drought stress environment, the distance between clusters ranged from 2.34 between cluster 4 and 6 to 49.43 between cluster 5 and 6 (Table 5.9).

The mean of each cluster for each trait were calculated and traits like biomass per plot, pod number per plot and yield per plot contributed a lot for distinct clusters at without drought stress environment and harvest index, yield per plot and biomass per plot also contributed for the observed cluster groups at drought stressed environment (Table 5.10).

The genotypic data analysis of core set accessions using hierarchical clustering with Ward method revealed a dendrogram depicting the genetic relatedness and difference of core set accessions. This analysis classified the 158 core set accessions in to eight distinct groups. The number of accessions clustered in each group ranged from 2 accessions in the 1^{st} group to 85 accessions in the 8^{th} group. Few genotypes were not grouped either of the groups which are very distinct from the other groups and each other (Table 5.11 & Fig. 5.3).

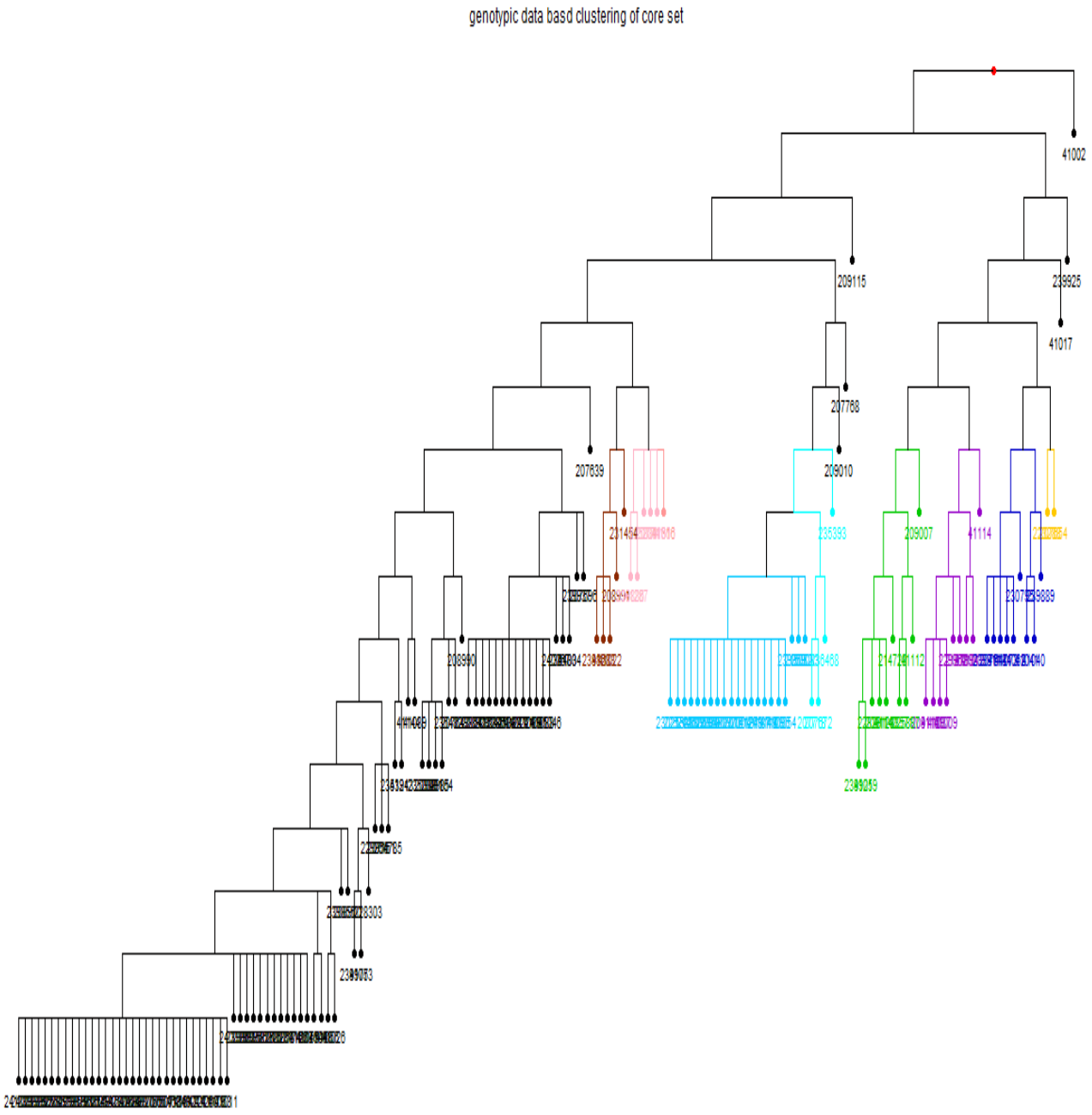

Fig. 5.3 Schematic representation of genotypic data based clustering of core set (yellow = cluster 1, blue = cluster 2, rose = cluster 3, green = cluster 4, light blue = cluster 5, light rose cluster 6, red = cluster 7 & black = cluster 8.

Core set Population Structure and Association Analysis

The Population structure of core set analysis was conducted using genotypic data of 55 polymorphic SNP markers using STRUCTURE software 2.3.3 (Pritchard *et al* 2000), and revealed that the accessions sub-divided into eight subpopulations. In these eight subpopulations, cluster one contained 24 accessions which accounted 15.19 % of the total core set accession membership, cluster two contained 6 accessions which accounted 3.79 % of the total core set membership, cluster three contained 17 accessions which contributed 10.76 % of the total core set membership, cluster four contained 27 accessions which contributed 17.08 % of the total core set membership, cluster five contained 49 accessions which contributed 31.01 % of the total core set membership, cluster six contained 14 accessions which contributed 8.86 % of the total core set membership, cluster seven contained one distinct accession and cluster eight contained 20 accessions which contributed 12.65 % of the total core set membership (Table 5.6 and Fig. 5.5).

The genotypic data based cluster distance analysis revealed the highest between cluster distance at cluster seven and cluster one (0.276) and the lowest between cluster distance at cluster five and cluster four (0.018). The average distance or expected heterozygosity between individuals in the same cluster revealed that cluster one showed the highest distance (0.0526) and cluster eight showed the lowest distance (0.0096) within the group.

The population relation differentiation (F_{ST} value) of each cluster mean ranged from 0.6419 (cluster one) to 0.9141(cluster eight). According to Odong *et al*. (2011), weak population differentiation showed lower F_{ST} value (F_{ST} <0.05) and high level differentiation showed higher F_{ST} value (F_{ST} >0.2). So, the analysis of population differentiation revealed high level of subpopulations fixation (F_{ST} >0.2) in each cluster (Table 5.4).

The gene diversity (expected heterozygosity) ranged from 0.0126 (CKaM0723) to 0.5062 (CKaM0033) with average gene diversity of 0.0710 and maximum polymorphic information content of 0.3843 CKaM0033) with low heterozygosity (average of 0.0188). From 55 polymorphic SNP markers, 9 showed heterozygosity with the highest heterozygosity was observed in CKaM0033 (0.9810).The result also showed that 17262 total alleles were detected. The average frequency of major allele is 0.9543 with range of 1(CKaM0025) to 0.6329 (CKaM1293) and the amount of heterozygosity is ranged from 0.9810 (CKaM0033) to 0.0063 (CKaM1101) (Table 5). The observed common alleles were 381 and rare alleles were 29 in the constructed core set population.

The analysis of marker trait association result revealed that there was strong association between markers and agronomically important traits at without drought stress environment in the general linear model (GLM) and mixed linear model (MLM) program analysis.

In general linear model analysis, a total of three markers were highly associated with agronomic traits. Marker CKaM0804, CKaM0993 and CKaM1902 were strongly associated with seeds number per plant, days to maturity and plant height, respectively. There was also highly significant association between markers and agronomic traits in mixed linear model (MLM) analysis. In this analysis, a total three markers showed strong association with agronomic traits. CKaM0804 strongly associated with seeds number per plant, CKaM1902 strongly associated with plant height and CKaM0993 strongly associated with days to maturity and plant height at without drought stress environment (Table 5.2).

Table 5.2 List of markers associated with agronomic traits at without drought stress environment in mixed linear (MLM) & general linear model (GLM) analysis.

Sr no.	Marker name	Associated trait	Environment	Model used
1	CKaM0804	SPP**	WOD	MLM
2	CKaM1902	PHT**	WOD	MLM
3	CKaM0993	DM**, PHT**	WOD	MLM
4	CKaM0804	SPP**	WOD	GLM
5	CKaM0993	DM**	WOD	GLM
6	CKaM1902	PHT**	WOD	GLM

** = significant at (P< 0.01), SPP = seeds per plant, PHT = plant height, DM = days to maturity, WOD = without drought stress environment

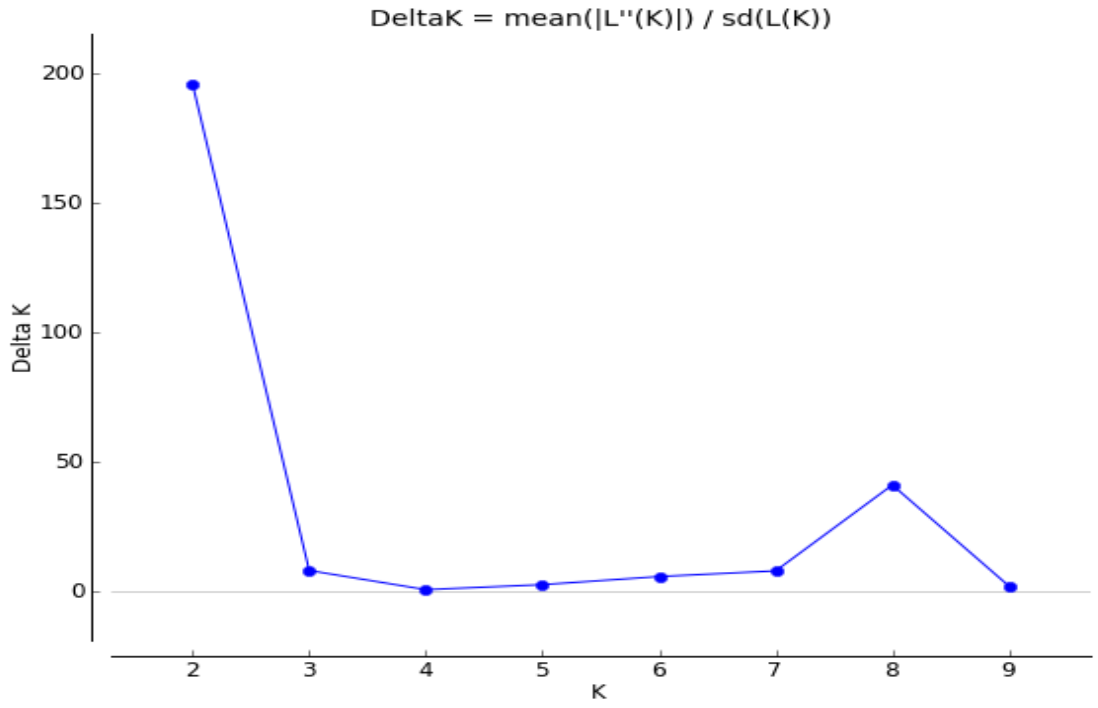

Fig 5.4 Plotting Delta K value to estimate the correct cluster number (K= 8)

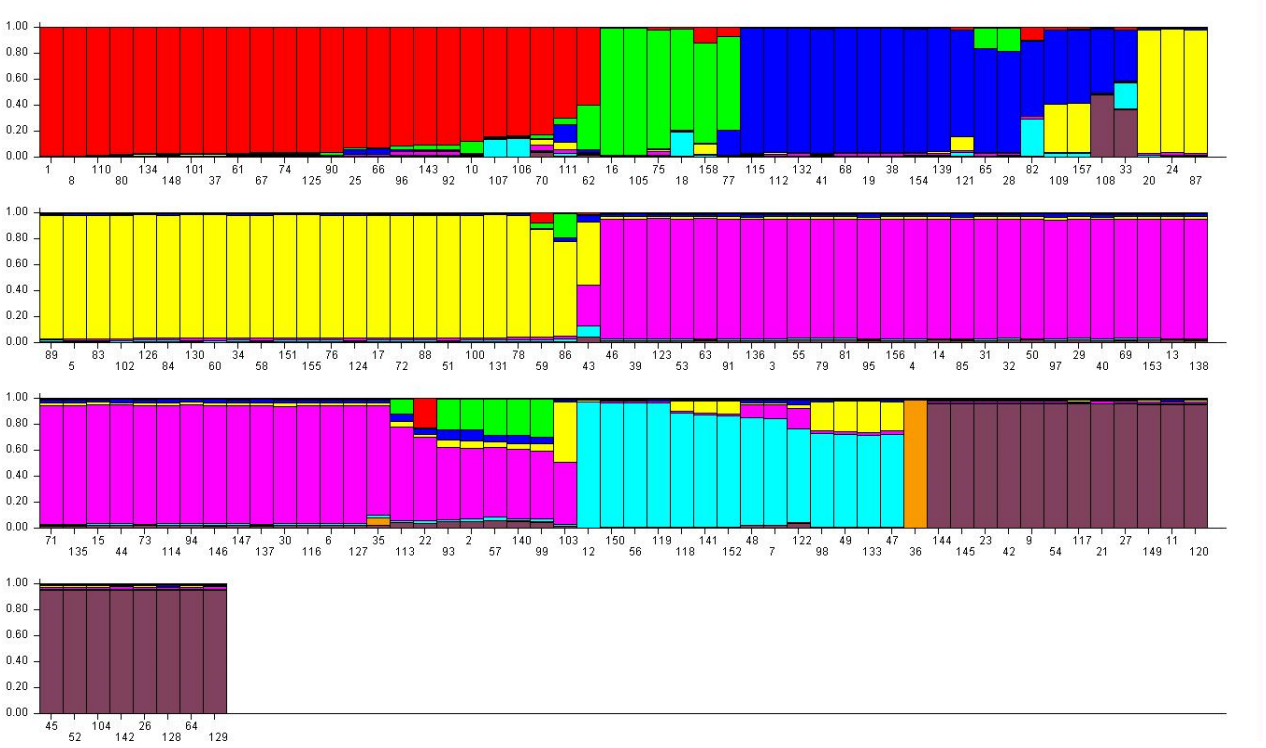

Fig 5.5 Population structure of core set, each line represents each accession with K color of cluster

Table 5.3 Number of genotyped accessions in each geographical regions and number of accessions selected for core set from each group

Sr No.	Regions	Number of accessions	Number of cluster	No. of accessions selected for core set (15 %)
1	Amara	451	16	69
2	Oromiya	284	6	43
3	SNNP	45	3	7
4	Tigray(including Eritra collections)	65	4	10
5	Unknown	151	9	25
6	Somalia*	3	-	2
7	Gambela*	1	-	1
8	Benshangul Gumz*	2	-	1
	Total	1002		158

* = Not considered percentage, SNNP = Southern Nations and Nationalities People

Table 5.4 Distribution of accessions in each cluster group based on phenotypic data at non- drought stress environment

Clusters	List of accessions in each cluster
I(2)	41002, 41009
II (39)	207700, 207696, 239905, 41140, 236482, 207672, 227153, 208977, 239863, 207151, 207647, 209114, 216856, 236883, 41279, 41098, 239886, 204785, 41112, 214730, 228793, 239911, 236193, 202509, 208990, 41263, 212474, 208991, 239855, 207768, 236480, 41283, 41017, 209094, 207741, 209091, 41103, 41099, 225878
III (31)	41011, 41026, 41287, 41134, 209090, 207666, 239889, 41046, 41169, 207718, 41054, 207174, 229962, 41322, 229956, 212476, 235033, 223288, 41034, 236491, 41271, 207163, 214916, 41062, 208993, 207607, 208829, 41312, 209083, 207644, 208900
IV(27)	207156, 213051, 215188, 235393, 41073, 229960, 207701, 41319, 41039, 207166, 41149, 239925, 228900, 41019, 219799, 208985, 228291, 207152, 207639, 41095, 241803, 41063, 239977, 239932, 207191, 225742, 209007
V (59)	41213, 41041, 41096, 41114, 239915, 240077, 239937, 41187, 41144, 41194, 215032, 214622, 219800, 41022, 225887, 41037, 209115, 207612, 233571, 225880, 230253, 41053, 41274, 210858, 230795, 236468, 41055, 41056, 41277, 229955, 240078, 207642, 225875, 214728, 225873, 209010, 231454, 41164, 222863, 239976, 41316, 239959, 228303, 239864, 207707, 215033, 239253, 41179, 41224, 223143, 41088, 207654, 41040, 209110, 239960, 41260, 240067, 41219, 235394

Number in parentheses indicate accession number in each cluster

Table 5.5 Distribution of accessions in each cluster group based on phenotypic data at drought stress environment

Clusters	List of accessions in each cluster
I (16)	41019, 41039, 209090, 41017, 41037, 228900, 41041, 225878, 41140, 209083, 41034, 239932, 210858, 219799, 209091, 213051
II (27)	207639, 240067, 228793, 230253, 41274, 207156, 225880, 236480, 240077, 41179, 207768, 235033, 239863, 41022, 208991, 204785, 239864, 41073, 209110, 41095, 209094, 215032, 216856, 41063, 207672, 41056, 41187
III (16)	239977, 41219, 239960, 207642, 207700, 207607, 228303, 207191, 239976, 207612, 225875, 41040, 41099, 41112, 208977, 225742
IV (35)	41088, 239915, 214916, 233571, 209010, 223288, 202509, 208985, 229960, 208993, 236468, 239889, 225873, 41263, 235393, 215188, 240078, 207654, 207741, 209007, 223143, 41096, 225887, 207707, 229962, 41098, 239925, 208900, 41114, 41194, 41224, 229955, 239905, 207647, 231454
V (2)	41054, 41062
VI (59)	41103, 212474, 236491, 214730, 236883, 235394, 239959, 229956, 207701, 208990, 209114, 41319, 236193, 207644, 208829, 41283, 207152, 227153, 41260, 239253, 207174, 239855, 239937, 41277, 41279, 41287, 207163, 41312, 207151, 212476, 41055, 41134, 41026, 41213, 41169, 214622, 222863, 41271, 214728, 207666, 239911, 41144, 41164, 219800, 228291, 207166, 207696, 41053, 241803, 215033, 236482, 239886, 41149, 41316, 207718, 41322, 41046, 209115, 230795
Out of cluster	41002, 41009, 41011

Number in parenthesis indicate accession number in each cluster

Table 5.6 Summary of cluster distance, Fst value and proportion of membership of core set

Cluster	1	2	3	4	5	6	7	8	Proportion of member ship	Mean value of Fst
1	**0.0526**								0.145	0.6419
2	0.0545	**0.0262**							0.053	0.8085
3	0.0716	0.076	**0.014**						0.101	0.8534
4	0.0598	0.0783	0.0348	**0.0099**					0.184	0.9047
5	0.0696	0.062	0.018	0.0181	**0.0097**				0.280	0.909
6	0.0769	0.0838	0.0408	0.0228	0.0243	**0.0344**			0.089	0.7636
7	0.276	0.2654	0.257	0.2276	0.2429	0.2217	**0.0924**		0.008	0.7754
8	0.0772	0.080	0.0359	0.0362	0.0181	0.0422	0.2599	**0.0096**	0.139	0.9141

Diagonal bold values are within cluster distance (expected heterozygosity b/n individual)

Table 5.7 Summary of major allele frequency, gene diversity, heterozygosity and polymorphic information content (PIC) value of core set

Sr no.	Markers name	Major allele frequency	Gene diversity	Heterozygosity	PIC value
1	CKaM0003	0.9937	0.0126	0.0000	0.0125
2	CKaM0017	0.9873	0.0250	0.0000	0.0247
3	CKaM0025	1.0000	0.0000	0.0000	0.0000
4	CKaM0033	0.5032	0.5062	0.9810	0.3843
5	CKaM0042	0.9937	0.0126	0.0000	0.0125
6	CKaM0043	1.0000	0.0000	0.0000	0.0000
7	CKaM0167	0.9873	0.0251	0.0000	0.0249
8	CKaM0204	0.9937	0.0126	0.0000	0.0125
9	CKaM0234	1.0000	0.0000	0.0000	0.0000
10	CKaM0249	0.9937	0.0126	0.0000	0.0125
11	CKaM0290	1.0000	0.0000	0.0000	0.0000
12	CKaM0317	0.8133	0.3060	0.0063	0.2629
13	CKaM0321	1.0000	0.0000	0.0000	0.0000
14	CKaM0343	0.9937	0.0126	0.0000	0.0125
15	CKaM0405	0.9810	0.0373	0.0000	0.0366
16	CKaM0411	0.9937	0.0126	0.0000	0.0125
17	CKaM0447	1.0000	0.0000	0.0000	0.0000
18	CKaM0462	0.8576	0.2460	0.0063	0.2187
19	CKaM0477	0.9873	0.0251	0.0000	0.0249
20	CKaM0493	1.0000	0.0000	0.0000	0.0000
21	CKaM0526	1.0000	0.0000	0.0000	0.0000
22	CKaM0588	1.0000	0.0000	0.0000	0.0000
23	CKaM0612	0.9873	0.0250	0.0000	0.0247
24	CKaM0630	0.9937	0.0126	0.0000	0.0125
25	CKaM0639	0.9873	0.0250	0.0000	0.0247
26	CKaM0647	0.9937	0.0126	0.0000	0.0125
27	CKaM0657	0.9684	0.0618	0.0000	0.0608
28	CKaM0707	1.0000	0.0000	0.0000	0.0000
29	CKaM0722	0.9937	0.0126	0.0000	0.0125
30	CKaM0723	0.9937	0.0126	0.0000	0.0125
31	CKaM0750	0.9684	0.0616	0.0000	0.0603

32	CKaM0804	0.8987	0.1843	0.0000	0.1713
33	CKaM0888	0.9810	0.0374	0.0000	0.0370
34	CKaM0993	0.9399	0.1142	0.0063	0.1100
35	CKaM0999	0.9430	0.1074	0.0000	0.1017
36	CKaM1101	0.8133	0.3060	0.0063	0.2629
37	CKaM1140	0.9557	0.0852	0.0000	0.0825
38	CKaM1175	0.9747	0.0496	0.0000	0.0488
39	CKaM1190	1.0000	0.0000	0.0000	0.0000
40	CKaM1254	1.0000	0.0000	0.0000	0.0000
41	CKaM1293	0.6329	0.4736	0.0127	0.3728
42	CKaM1317	0.9937	0.0126	0.0000	0.0125
43	CKaM1328	0.9937	0.0126	0.0000	0.0125
44	CKaM1356	0.9937	0.0126	0.0000	0.0125
45	CKaM1641	0.9810	0.0373	0.0000	0.0366
46	CKaM1651	0.7943	0.3339	0.0063	0.2894
47	CKaM1788	0.9747	0.0496	0.0000	0.0488
48	CKaM1832	0.9937	0.0126	0.0000	0.0125
49	CKaM1842	0.9873	0.0251	0.0000	0.0249
50	CKaM1848	0.9684	0.0616	0.0000	0.0603
51	CKaM1902	0.9114	0.1615	0.0000	0.1485
52	CKaM1903	0.8165	0.3020	0.0000	0.2600
53	CKaM1933	1.0000	0.0000	0.0000	0.0000
54	CKaM1963	0.9810	0.0373	0.0000	0.0366
55	CKaM1971	0.9905	0.0188	0.0063	0.0187
	Mean	0.9543	0.0710	0.0188	0.0624

Table 5.8 Distribution of accessions in each cluster (clustering accessions using STRUCTURE software)

Cluster	List of accessions
I (24)	207672, 207607, 239886, 41039, 225742, 209114, 212476, 41099, 207644, 208977, 207191, 207647, 41194, 236468, 239925, 229962, 214916, 209094, 219799, 219800, 207707, 41002, 41034, 223288
II (6)	207718, 41062, 241803, 207768, 41055, 216856
III (17)	222863, 41164, 239253, 228900, 239977, 225880, 41063, 207654, 208990, 41260, 207642, 41213, 235394, 41114, 225873, 240078, 223143
IV (27)	41271, 207741, 41056, 207166, 209090, 208829, 235393, 208993, 240067,214730, 41169, 209091, 41319, 41098, 229960, 239959, 230253, 207163, 207700, 209083, 208991, 41019, 235033, 209010, 207174, 215032, 41073
V (49)	207156, 239855, 41009, 214728, 209115, 225875, 41095, 41022, 225887, 210858, 230795, 208900, 41054, 207151, 41179, 41134, 225878, 41046, 207701, 41149, 207666, 41274, 207612, 41140, 41053, 208985, 41017, 41144, 41224, 239976, 239915, 236883, 236480, 212474, 209007, 41316, 207696, 239911, 236491, 236482, 209110, 240077, 229956, 41011, 41279, 202509, 213051, 41219, 215033
VI (14)	229955, 41026, 41287, 214622, 41283, 41312, 228291, 236193, 239863, 207152, 228303, 239937, 239960, 41041
VII	41187
VIII (20)	239889, 239905, 41263, 41096, 204785, 41037, 227153, 41112, 41088, 239932, 41040, 228793, 239864, 41277, 215188, 41322, 231454, 41103, 207639, 233571

Numbers in parentheses indicated number of accessions in each cluster group

Table 5.9 Drought tolerance indices of selected best accessions from the core set of chickpea.

Accessions ID	WDS yield	WODS yield	RP	DTE %	DSI	MP	TDS
41034	392.75	210.50	1.87	186.58	-1.019	301.63	182.25
41040	380.80	245.65	1.55	155.02	-0.647	313.23	135.15
41011	374.75	240.00	1.56	156.15	-0.661	307.38	134.75
215033	355.40	222.45	1.60	159.77	-0.703	288.93	132.95
41026	391.30	280.25	1.40	139.63	-0.466	335.78	111.05
239932	364.85	258.70	1.41	141.03	-0.483	311.78	106.15
228900	347.25	243.55	1.43	142.58	-0.501	295.40	103.70
41073	328.80	229.70	1.43	143.14	-0.508	279.25	99.10
207612	291.25	211.15	1.38	137.94	-0.446	251.20	80.10
41039	348.25	268.45	1.30	129.73	-0.350	308.35	79.80
207647	279.75	200.15	1.40	139.77	-0.468	239.95	79.60
41002	558.50	489.00	1.14	114.21	-0.167	523.75	69.50
207152	253.45	190.50	1.33	133.04	-0.389	221.98	62.95
207696	305.30	251.05	1.22	121.61	-0.254	278.18	54.25
41017	329.05	281.55	1.17	116.87	-0.198	305.30	47.50
208977	256.50	210.25	1.22	122.00	-0.259	233.38	46.25
41063	305.55	259.50	1.18	117.75	-0.209	282.53	46.05
215032	353.20	317.10	1.11	111.38	-0.134	335.15	36.10
41096	304.80	270.05	1.13	112.87	-0.151	287.43	34.75
236468	273.55	241.75	1.13	113.15	-0.155	257.65	31.80
41263	294.00	262.25	1.12	112.11	-0.142	278.13	31.75
41144	365.00	334.85	1.09	109.00	-0.106	349.93	30.15
214730	336.60	314.85	1.07	106.91	-0.081	325.73	21.75
225878	313.65	292.20	1.07	107.34	-0.086	302.93	21.45
41164	272.95	252.00	1.08	108.31	-0.098	262.48	20.95
223143	307.45	287.80	1.07	106.83	-0.080	297.63	19.65
235393	302.55	284.45	1.06	106.36	-0.075	293.50	18.10
41009	387.50	370.55	1.05	104.57	-0.054	379.03	16.95
240077	304.10	290.40	1.05	104.72	-0.056	297.25	13.70
41312	294.80	282.15	1.04	104.48	-0.053	288.48	12.65
Mean	332.45	269.76	1.25	125.49	-0.299	301.11	62.69
Kutaye	265.50	230.00	1.15	115.43	-0.182	247.75	35.50
Fetenech	265.25	332.05	0.7988	79.883	0.237	298.65	-66.80
Mariye	237.80	281.35	0.8452	84.521	0.182	259.58	-43.55

Where WDS= yield at drought stressed environment, WODS= yield at without drought stress environment, RP = rate of productivity, DTE = Drought tolerance efficiency, DSI = drought susceptibility index, MP = mean productivity, TDS = Tolerant to drought stress.

Table 5.10 Summary of core set validation indices for each evaluated traits at two environments

Trait	Without Drought stress environment				With drought stress environment			
	MD %	VD %	CR %	VR %	MD %	VD %	CR %	VR %
BM	1.486	8.696	86.809	93.94	3.091	25.000	91.765	109.091
DF	0.090	2.394	88.889	100.00	0.346	9.823	88.889	100.000
HI	0.847	0.000	95.349	100.00	0.467	3.704	100.000	104.167
HSW	10.789	13.472	46.154	50.00	0.073	15.436	100.000	112.500
DM	0.154	2.030	100.000	100.00	0.010	1.919	100.000	100.000
PFP	0.234	4.867	86.957	100.00	0.368	5.923	94.737	100.000
PHT	1.080	1.708	42.553	100.00	0.348	0.568	96.875	100.000
PPP	1.908	4.178	91.429	105.00	4.283	11.094	55.046	103.571
SPP	0.640	0.507	62.626	100.00	0.243	4.429	100.000	100.000
YLD	0.546	3.324	97.233	100.00	2.214	37.494	100.000	110.000
YPP	10.488	32.161	208.167	200.00	2.074	6.667	81.221	94.444

MD % = percent mean difference, VD % = percent variance difference, CR % = Coincidence rate of range, VR %= Variable rate of coefficient of variance, BM = biomass per plot, DF= Days 50 % flowering, HI = Harvest index, HSW= hundred seeds weight, DM = maturity days, PFP = pod filling period, PHT= plant height, PPP = pod per plant, SPP= seed per plant, YLD = yield per plot, YPP= yield per plant.

Table 5.11 Summary of between clusters distance of core set based on phenotypic data for the evaluated traits at two environments

Cluster	I	II	III	IV		Environment
II	21.81	-				at without drought stress environment
III	20.12	3.51	-			
IV	19.46	3.71	3.44	-		
V	21.02	3.41	3.72	4.03		
Cluster	I	II	III	IV	V	at drought stressed environment
II	5.19	-				
III	10.91	8.80	-			
IV	5.77	3.59	14.24	-		
V	46.50	43.66	43.69	42.21		
VI	5.24	3.43	14.22	2.34	49.43	

Table 5.12 Summary of cluster mean of core set at without drought stress environment

Trait/Cluster	Cluster means				
	I	II	III	IV	V
BM	1.85	1.36	1.51	1.59	1.54
DF	51	52.02	52.11	52.14	52.6
HI	0.25	0.25	0.22	0.22	0.23
HSW	15.75	14.74	14.98	15.26	15.13
DM	96.5	94.86	95.06	95.08	95.07
PFP	45.5	42.81	42.95	42.94	42.47
PHT	47.0	47.03	47.16	47.16	46.77
PPP	56.5	58.46	54.26	56.66	56.4
YLD	429.78	322.29	316.18	329.41	324.84
YPP	16.02	13.58	13.45	13.64	13.39
SPP	75.5	76.16	74.81	77.06	75.51

BM = Biomass per plot, DF = Days to 50 % flowering, HI = Harvest index, HSW = Hundred seeds weight, DM = Days to maturity, PFP = Pod filling period, PHT = Plant height, PPP = Pod per plant, YLD = Yield per plot, YPP = Yield per plant, SPP = Seed number per plant.

Table 5.13 Distribution of accessions in each cluster based on SNP marker data clustering

Cluster	List of accessions
I (2)	223288, 207654
II (9)	239889, 41040, 41041, 230795, 41062, 41073, 41134, 207647, 239911
III (9)	41114, 41019, 41194, 219799, 229962, 41009, 41037, 41169, 209110,
IV (10)	209007, 41112, 225887, 233571, 214728, 41140, 209114, 228291, 41219, 239905
V (25)	235393, 236468, 207672, 207718, 209083, 215032, 239959, 41054, 41096, 41103, 41149, 41187, 41279, 207191, 207612, 207700, 207701, 208829, 209090, 209094, 213051, 214622, 215188, 225742, 230253
VI (6)	41316, 207166, 207741, 225880, 41287, 239886,
VII (5)	231454, 208991, 41022, 41283, 239863
VIII (85)	207696, 239937, 41034, 236480, 240067, 41046, 41088, 41095, 41099, 41224, 41277, 41322, 202509, 222863, 223143, 227153, 228900, 239864, 208990, 207642, 236482, 41164, 208985, 209091, 235033, 41026, 41056, 41098, 207644, 41055, 41263, 207152, 207174, 210858, 225875, 225878, 225873, 229960, 239915, 239960, 240078, 41039, 41144, 41213, 235394, 204785, 207151, 229955, 228303, 41053, 239977, 208900, 239855, 41011, 41063, 41179, 41260, 41271, 41274, 41312, 41319, 207156, 207163, 207607, 207666, 207707, 208977, 208993, 212474, 212476, 214730, 214916, 215033, 216856, 219800, 228793, 229956, 236193, 236491, 236883, 239253, 239932, 239976, 240077, 241803
Out of any group	41002, 239925, 41017, 209010, 207768, 209115, 207639

Number in parentheses indicate number of accessions in each cluster

Table 5.14 Mean, standard deviation, range and coefficient of variance for core set and entire collection at drought stressed environment

Trait	Entire set				Core set			
	Mean	Stan. deviation	range	Coefficient of variance	Mean	Stan. deviation	range	Coefficient of variance
BM	1.29	0.28	1.70	22.00	1.33	0.32	1.56	24.00
DF	49.14	3.53	18.00	7.00	49.31	3.35	16.00	7.00
HI	0.21	0.052	0.33	24.00	0.21	0.05	0.33	25.00
HSW	15.15	1.22	7.00	8.00	15.14	1.31	7.00	9.00
DM	92.91	2.17	13.00	2.00	92.92	2.19	13.00	2.00
PFP	43.77	3.39	19.00	8.00	43.61	3.29	18.00	8.00
PHT	45.66	5.31	32.00	12.00	45.82	5.33	31.00	12.00
PPP	62.17	17.59	218.00	28.00	64.83	18.57	120.00	29.00
SPP	76.67	13.81	72.00	18.00	76.48	13.52	72.00	18.00
YLD	268.84	52.78	401.10	20.00	274.79	61.97	401.10	22.00
YPP	14.46	2.59	17.04	18.00	14.76	2.51	13.84	17.00

Table 5.15 Pearson's correlation of eleven agronomic traits of core set (without drought stress environment).

	DF	DM	BM	YLD	HSW	HI	PFP	PPP	YPP	PHT
DM	0.4083**									
BM	0.0037ns	0.027ns								
YLD	0.0497ns	0.0307ns	0.2862**							
HSW	0.0324ns	0.0235ns	0.2572**	0.5101**						
HI	0.0368ns	-0.0005ns	-0.6918**	0.4414**	0.1293**					
PFP	-0.4727ns	0.5908**	0.0154ns	-0.0126ns	-0.0019ns	-0.0273ns				
PPP	0.0228ns	0.0123ns	0.2275**	0.2739**	0.2757**	-0.017ns	-0.0172ns			
YPP	0.032ns	0.0194**	0.1389ns	0.1964**	0.1844**	-0.0003ns	-0.0135ns	0.4175**		
PHT	0.001ns	0.0899**	0.1061**	0.1505**	0.1633**	-0.001ns	0.0762*	0.0873**	0.1256**	
SPP	0.0492ns	0.0541ns	0.2308**	0.5005**	0.3658ns	0.1595ns	0.0085ns	0.382**	0.2237ns	0.1264**

BM = biomass per plot, DF = days to 50 % flowering, HI = harvest index, HSW = hundred seeds weight, DM = days to maturity, PFP = pod filling period, PHT = Plant height, PPP = pod per plant, SPP = seeds number per plant, YLD = yield per plot, YPP = yield per plant, * = significant at ($p < 0.05$) and ** = significant at ($p < 0.01$)

Table 5.16 Pearson's correlation of eleven agronomic traits of core set (at drought stressed environment)

	DF	DM	PHT	BM	YLD	HSW	HI	PFP	PPP	SPP
DM	0.3449**									
PHT	0.1066**	0.1819**								
BM	-0.0639**	-0.0053ns	-0.0087ns							
YLD	0.0853**	-0.0266ns	-0.0148ns	0.2367**						
HSW	0.151**	0.0409ns	0.0489*	0.1127**	0.6028**					
HI	0.1154**	-0.0065ns	0.0125ns	-0.4875**	0.5394**	0.3517**				
PFP	-0.8192**	0.2557**	0.0013ns	0.0626**	-0.1041**	-0.1306**	-0.1229**			
PPP	-0.0683**	-0.0134ns	0.0141	0.1995**	0.1601**	0.0992**	-0.002ns	0.0622**		
SPP	0.0843**	0.0617**	0.0449**	0.0819**	0.2639**	0.2599**	0.1513**	-0.0491*	0.0784**	
YPP	-0.0155ns	0.0058**	-0.0053ns	0.1959**	0.2616**	0.1677**	0.0775**	0.0195ns	0.4252**	0.0663**

BM = biomass per plot, DF = days to 50 % flowering, HI = harvest index, HSW = hundred seeds weight, DM = days to maturity, PFP = pod filling period, PHT = Plant height, PPP = pod per plant, SPP = seeds number per plant, YLD = yield per plot, YPP = yield per plant, * = significant at ($p< 0.05$) and ** = significant at ($p< 0.01$)

Table 5.17 Summary of Shannon-Weaver diversity index (H) for core set and entire collection

Traits	Core set		Entire collection	
	WDS	WODS	WDS	WODS
DF	1.21	1.09	1.44	1.28
DM	1.21	0.87	1.33	1.28
PHT	0.99	1.24	1.30	1.19
PPP	1.39	0.88	1.32	1.24
SPP	1.02	1.24	1.49	1.29
BM	0.86	1.33	1.25	1.45
YPP	1.40	0.82	1.48	1.22
HSW	0.82	1.32	0.98	1.33
YLD	1.29	1.13	1.41	1.42
HI	1.55	1.01	1.40	1.12
PFP	1.08	1.17	1.42	1.35

DF = days to 50 % flowering, DM = days to maturity, PHT = plant height, PPP = pod per plant, SPP = seed number per plant, BM = biomass per plot, PHT = plant height, PPP = Pod per plant, SPP = seed number per plant, YPP = yield per plant, HSW = hundred seeds weight, YLD = yield per plot, HI = harvest index and PFP = pod filling period

5.4. Discussion

Germplasm collections were originally set up to preserve the genetic diversity of crop species and their wild relatives. Given that such genetic diversity of crops has an economic value, conservation for use has been the driving force behind many gene banks. The sheer number of accessions making up germplasm collections could be an obstacle for their full exploitation, evaluation and utilization to impact the crop improvement or breeding programmes.

In this regard, genetic diversity of such a large collection may not have been adequately evaluated for various biotic and abiotic stresses, due to resource and time constraints. It is impractical to evaluate such large collection in detail as it would be expensive and time consuming. This task could be more easily fulfilled by developing subsets of the whole collection, called active working collections by Harlan (1972) and core collection by Frankel and Brown (1984).

In the present study, chickpea germplasm core set was developed based on passport data SNP marker data. Based on passport data (collection site description), accessions were stratified in to five major geographical origin groups. Cluster analysis of each group based on SNP marker data revealed distinct cluster groups based on relatedness and genetic similarity. Accessions were drawn randomly from each similar accession from each cluster group using 15 % sampling intensity proportionally. The constructed 158 core set accessions were validated for different core set validation parameters such percentage variance difference (VD %), percentage mean difference (MD %), the coincidence rate of range (CR %), variable rate of coefficient of variance (VR %) and Shannon-Weaver diversity index. The validation index revealed better correspondence between entire collection and core set collection.

In addition to this, the population structure and allelic diversity of the entire collection and core set collection was significantly similar in population differentiation and fixation, gene diversity, average of major allele frequency, polymorphic information content value and proportion of detected total, common and rare alleles. Marker – trait association analysis also revealed that 14 markers showed highly significant and significant association with the evaluated trait at both environments. The cores set correctly represent the diversity of entire collection of Ethiopian chickpea germplasm with the reduction of germplasm redundancy or duplication.

Desirable accessions were also identified based on yield stability (drought susceptibility index, drought tolerance efficiency and tolerance to drought stress) across drought stressed and without drought stressed environment. Based on higher value of drought tolerance efficiency, tolerance to drought stress and lower value of drought susceptibility index, accessions showed better performance than checks (Kutaye and Fetenech) released for drought prone environments. Similar findings were reported by (Upadhyaya *et al*. 2008) in the global core collection development and evaluation.

The core collection will be also important point of entry to further research and the proper exploitation of the genetic resources available in Ethiopia by reducing time and money to screen the whole germplasm from gene bank.

6. Summary and Conclusions

The present finding entitled "Development and Utilization of Genetic Diversity Based Ethiopian Chickpea (*Cicer arietinum* L.) Germplasm Core Collection for Association Mapping" was conducted with the following objectives.

1. Preliminary phenotyping and genotyping of germplasms
2. Development of chickpea core collection based on diversity analysis.
3. Identification of desirable accessions for drought tolerance from core set by proper phenotyping
4. Large scale genotyping of the core collections by SNP markers
5. Quantification of Linkage Disequilibrium using molecular marker data
6. Identification and establishing marker trait associations using appropriate association genetic approaches.
7. Quantification of population structure and relationship of Ethiopian chickpea collection

The findings from this research work are briefly summarized below

1. Germplasm characterization and diversity assessment is the first and most important criteria to utilize in the breeding program and manage the gene bank collections. Based on the phenotypic data and SNP marker data, almost the total chickpea collections characterized. The hierarchical diversity analysis revealed that there is more diverse germplasms that to be exploited for their desirable traits especially the germplasm which were collected in the earlier time showed better genetic diversity than the one collected in the recent time. The analysis also revealed that the studied traits showed strong correlation with yield and each other indicating that simultaneous selection can improve other desirable associated traits. The

principal component analysis also revealed that five main components contributed 75-76.6 % of the total phenotypic variance at two contrasting environments.

2. The population structure and relationship analysis revealed six distinct sup populations were identified and the subpopulations showed strong differentiation from the original population (Fst> 0.2). Admixtured and migrant individuals were identified in each group using STRUCTURE software. The genetic distance between all of these groups showed better diversity. The gene diversity or expected heterozygosity showed better gene diversity and more than 109 thousand alleles detected with better common and rare alleles in the total population. It showed that there are diverse alleles in the germplasms which could be further assessed for interest of traits.
3. The linkage disequilibrium (LD) analysis revealed strong LD among markers indicating there is sufficient prospect for association mapping even if marker numbers are sparsely covered the linkage groups.
4. The trait- marker association analysis revealed strong marker- trait associations on maturity related traits and on biomass per plot. It is very important for indirect selection for drought resistance (drought escape) traits or genes.
5. To address the problem of huge germplasm utilization present in the gene bank, the main gateway is construction of representative core collection. National based chickpea core collection was constructed based on proper sampling procedures and validation parameters with 15 % sampling intensity.

   The core collection was constructed using stratified random sampling procedure expecting to retain 85 % or above genetic diversity representation of the original population with reduction or totally excluding the duplicated accessions in the core set. Based these

procedure and sampling intensity, 158 diverse accessions that represent the country chickpea growing area and genetic variability, are identified which are useful for developing varieties though direct selection or used as parent materials in the crossing program.

6. The constructed core set or accessions were validated using standardized validation parameters. Based on the phenotypic data, accessions were evaluated for representative indices. The indices were the average of absolute difference between means and variances across all the traits for core set and entire set. The goodness of indices showed better correspondence of core set and entire set. Core set also evaluated for percent of coincidence rate (CR %) and percent variable rate (VR %) and the result showed better representation. Shannon-Weaver diversity index (H) and analysis of variance indicated there is no significant difference between core set and entire collection for the evaluated eleven quantitative traits. This showed that the developed core set representing correctly the Ethiopian chickpea germplasm entire collection without any accession duplication in the core set.

7. As terminal drought is the most chickpea production limiting factor in Ethiopia and other chickpea growing countries, identifying desirable genotypes for drought resistance from the developed core set is worth full activity for plant breeders. So based on drought tolerance efficiency, drought susceptibility index, tolerance to drought stress, rate of productivity and mean productivity, around 30 accessions were identified as drought tolerance and better than the check varieties which are released for drought resistance for the country.

8. As a future direction, the developed core set has to be utilized by the national chickpea improvement program and the identified drought tolerant accessions from the core set has to be further checked in better environment (green house).
9. Lastly, we conclude that SNP markers are cost effective marker technology which generates extremely high quality data needed for germplasm diversity study and association analysis.

7. Literature Cited

Adams, W.T., Bikes, D.S. and Erickson, V.J. 1992. Using genetic markers to measure gene flow and pollen dispersal in forest tree seed orchards. In Ecology and evolution of plant reproduction, edited by R. Waytt, Chapman and Hall, New York, PP 37-61.

Agrobase, 2011, AGROBASE Generation III. Users manual. Version 33, Agromix software Winning, MB, Canada.

Ahmad, F. and Slinkard, A. E. 1992. Genetic relationships in the genus *Cicer* L. as revealed by polyacrylamide gel electrophoresis of seed storage proteins. *Theor. Appl. Gene*. 84: 688-692.

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H., Hayden, M. J., Howes, N., Sharp, P., Vaughan, P., Rathmell, B., Huttner, E. and Kilian, A. 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor Appl Gene,* 113: 1409-1420.

Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P., and Nordborg, M. 2005. Genome-Wide Association Mapping in *Arabidopsis* Identifies Previously Known Flowering Time and Pathogen Resistance Genes. *PLOS Genetics* 1: 531-539.

Araus, J. L., Slafer, G. A., Reynolds, M. P. and Royo, C. 2002. Plant breeding and drought in C-3 cereals: what should we breed for? *Ann. Bot*. 89: 925–940.

Arumuganathan, K. and Earle E.D. (1991) Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter* 9: 208–218.

Azhaguvel, P. S., Vidya, D., Sharma, A. and Varshney, R. K. 2006. Methodological advancement in molecular markers to delimit the gene(s) for crop improvement. *In*: Texiera

da Silva (ed.) Floriculture, Ornamental, Plant Biotechnology Advances and Tropical Issues. P 460-462.

Barrett, B.A., and K.K. Kidwell. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Sci*. 38:1261–1271.

Bayush Tsegaye and Trygue Berg, 2007. Genetic erosion of Ethiopian tetraploid wheat landraces in eastern Shewa, Central Ethiopia. *Genetic Resource and Crop Evolution*, 54: pp. 715-726.

Bernatzky, R. and Tanksley, S. D. (1986). Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics.* 112: 887-898.

Blears, M. J., De Grandis, S. A., Lee, H. and Trevors, J. T. 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *J. ind. microbiol. biotechnol*. 21**:** 99-114.

Blum, A. 1988. Plant breeding for stress environments. (CRC press: Boca Raton, FL)

Brown, A. H. D., 1989a, Core collections: A practical approach to genetic resources management. *Genome,* 31: 818-824.

Brown, A.H.D. 1989. Core collections: A practical approach to genetic resources management. *Genome,* 31:818–824.

Brown, A. H. D. 1989b. The case for core collections. p. 136–155. *In* A.H.D. Brown *et al*. (ed.). The use of plant genetic resources. Cambridge Univ. Press, Cambridge, England.

Brown, A.H.D. 1995. The core collection at the crossroads. In: Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L., Morales, E.A.V., (Eds.). *Core Collections of Plant Genetic Resources*. John Wiley and Sons, UK, 3-19.

Brugmans, B., van der Hulst, R. G., Visser, R. G. F., Lindhout, P. and van Eck, H. J. 2003. A new and versatile method for the successful conversion of AFLP markers into simple single locus markers. *Nucleic Acids Res* 31**:** e55.

Buckler, E.S. and J.M. Thornsberry. 2002. Plant molecular diversity and applications to genomics. *Current opinion in plant biology*. 5: 107-111.

Caldwell K.S., Russell J, Langridge P, Powell W. 2006. Extreme population dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*, vol. 172, No. 1: 557-567.

Churchill, GD, Airely, C., Allayee, H., Angel, J.M. and Attie, A.D. 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature Genetics.* 36**:** 1133-1137.

Collins, F. S., Guyer, M. S. and Charkravarti, A. 1997. Variations on a theme: cataloging human DNA sequence variation. *Science*, 278: 1580-1581.

Creelman, R. A. and Zeevaart, J.A. D. 1985. Abscisic acid accumulation in spinach leaf slices in the presence of penetrating and non-penetration solutes. *Plant Physiol*. **77**:25–28.

Crossa, J., I.H. DeLacy, and S. Taba. 1995a. The use of multivariate methods in developing a core collection. p. 77–89. *In* T. Hodgkin et al. (ed.) Core collections of Plant genetic resources. John Wiley & Sons Inc., New York.

Dasgupta, T., M.O. Islam, P. Gayen and K.K. Sarkak. 1987. Genetic divergence in chickpea (*Cicer arietinum* L.). *Exp. Genet*. 3: 15-21.

Davis, J. E., Kubisiak, T. L. and Milgroom, M. G. (2005). Polymorphic sequence-characterized codominant loci in the chestnut blight fungus, *Cryphonectria parasitica*. *Molecular ecology* notes 5: 195-197.

D'hoop, B.B., Paulo, M.J., Kowitwanich, K., Senger, M., Visser, R.G.F., van Eck, H.J. and van Eeuwijk, F.A. 2010. Population structure and linkage disequilibrium unraveled in tetraploid potato. *Theor Appl Gene.* 121: 1151-1170.

Dwevedi, K.K. and Gaibriyal M. Lal, 2009. Assessment of genetic diversity of cultivated chickpea. *Asian Journal of Agricultural Sciences* 1(1): 7-8.

Duke, J.A. 1981. Handbook of legumes of world economic importance. Plenum Press, New York. p. 52-57.

ECuA (Ethiopian Customs Authority). 2008. Report on export commodity. Addis Ababa, Ethiopia.

Falconer, D.S. and Mackey. T.F.C. 1996. Introduction to quantitative genetics 4th edition long man, Essex, UK.

FAO. 1998. Food and Agriculture Organization of the United Nations Statistics. (http://faostat.fao.org), Rome, Italy.

FAO. 2006. Food and Agriculture Organization of the United Nations Statistics. (http://faostat.fao.org), Rome, Italy.

FAO. 2009. Food and Agriculture Organization of the United Nations Statistics. (http://faostat.fao.org), Rome, Italy.

Farquhar, G. D and Richards, R. A. 1984. Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. Australian Journal of Plant Physiology 11: 539–552.

Fischer, R.A., and Maurer, R. 1978. Drought Resistance in spring wheat cultivar .I. Grain yield response. *Aust. J. of. Agric. Res*., 29:897-917.

Fischer, K. S. and Wood, G. 1981. Breeding and selection for drought tolerance in tropical maize. In: Proc. Symp. on Principles and Methods in Crop Improvement for Drought Resistance with Emphasis on Rice, IRRI, Philippines, 23-25th May, 1981.

Franco, J., J. Crossa, J. Villasen˜ or, S. Taba, and S.A. Eberhart. 1998. Classifying genetic resources by categorical and continuous variables. Crop Sci. 38:1688–1696.

Franco, J., J. Crossa, J. Villasen˜ or, S. Taba, and S.A. Eberhart. 1999. A two-stage, three-way method for classifying genetic resources in multiple environments. *Crop Sci*. 39:259–267

Franco, J., and J. Crossa. 2002. The Modified Location Model for classifying genetic resources. I: Association between categorical and continuous variables. *Crop Sci*. 42:1719–1726.

Franco, J., Crossa, J., Warbburton, M.L., Taba, S. 2006. Sampling strategies for conserving maize diversity when forming core subsets using genetic markers, *Crop Sci.* 46: 854-864.

Flint-Garcia, S.A, Thornsberry, J. M. and Buckler, E.S. 2003. Structure of linkage disequilibrium in plants. *Ann Rev Plant Biol*, 54:357–374

Frankel, O.H. 1984. Genetic perspective of germplasm conservation. p. 161–170. *In* W. Arber *et al*. (ed.). Genetic Manipulations: Impact on Man and Society. Cambridge Univ. Press, Cambridge, England.

Franke, O. H. and Brown, A. H. D., 1984, Current plant genetic resources- a critical appraisal. In : V. L. Chopra, B. C. Joshi, R. P. Sharma and H. C. Bansal (eds). *Genetics: New Frontiers* Vol. VI. Oxfort and IBH Publ. Co., New Delhi, pp. 1-11.

Gale, M. D. and Devos, K. M. 1998. Comparative genetics in the grasses. Proc. *Natl. Acad. Sci.* USA. 95: 1971-1974.

Gaur P.M., Tripathi S, Gowda CLL, Ranga Rao GV, Sharma HC, Pande S and Sharma M. 2010. Chickpea seed production manual. International Crop Research Institute for the Semi-Arid Tropics, Patancheru, Andrapradesh, India 28 PP.

Golabadi M, Arzani A, Mirmohammadi Maibody SAM 2006. Assessment of drought tolerance in segregating population in durum wheat. *Afri. J. Agric. Res.*, 1(5): 162-171.

Goldstein, D. B. and Pollock, D. D. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. J. *Hered.* 88: 335-342.

Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 27: 857-871.

Gupta, P. K., Roy, J. K. and Prasad, M. 2001. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr Science*, 80**:** 524-535.

Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N. and Balyan, H. S. 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics*, 270**:** 315-323.

Hafid R.E., Smith DH, Karrou M, Samir K. 1998. Morphological attributes associated with early-season drought tolerance in spring durum wheat in a Mediterranean environment. *Euphytica*, Vol. 101, 3: 273-282.

Hannan RM, Kaiser WJ and Muehlbauer F.J. 1994. Development and Utilization of the USDA Chickpea Germplasm Core Collection. Agronomy Abstracts 1994. Madison, Wisconsin: *American Society of Agronomy*, p. 217.

Harlan J.R. 1972. Genetics of disaster. *Journal of Environmental Quality* 1: 212–215.

Hsiao, T. C., Acevedo, E., Fereres, E. and Henderson, D.W. 1976. Stress metabolism: Water stress, growth, and osmotic adjustment. Philos, Trans. R. soc. London, 273: 479-500.

Helentjaris, T., Slocum, M., Wright, S., Schaefer, A. and Nienhuis, J. 1986. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 72: 761-769.

Henry, R. J. 2001. Plant genotyping: the DNA fingerprinting of plants. CABI Publishing, Wallingford.

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton D, McKeigue P. 2003. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 72: 1492–1504.

Hu J., Zhu J, Xu H.M. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Gene.* 101:264–268

IPGRI (International Plant Genetic Resources Institute). 1993. Diversity for Development: The stratagem of the International Plant Genetic Resources Institution. International Plant Genetic Resources Institution. Rome.

Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 29**:** 10-25.

Janick, J. 1995. Plant breeding review. Vol 13. John Wiley and Sons, Inc. Now York **.**

Jeffreys, A. J. 1979. DNA sequence variants in the GA and B - globin genes of Man. *Cell*. 18: 1-10.

Jordan, W. R., Dougas, J. R. and Shouse, P. J. 1983. Strategies of crop improvement for drought-prone regions. *Agric. Water Manage*. 7: 281

Kashiwagi, J., Krishnamurthy, L., Crouch, J. H. and Serraj, R. 2006. Variability of root length density and its contributions to seed yield in chickpea (*Cicer arietinum* L.) under terminal drought stress. *Field Crops Res*. 95:171-181.

Key, D. E., 1979. Food Legumes. Tropical Products Institute. London, PP. 48-71.

Khanna-Chopra, R. and Sinha, S. K. 1987. Chickpea: Physiological aspects on growth and yield. In: M.C. Saxena & K.B. Singh (Eds.). The Chickpea, pp. 163–189. CAB International, Wallingford, UK

Kim, K.W. Chung, H.K., Cho, G.T., Ma, K.H., Chandrabalan, D. Gwag, J.G., Kim, T.S., Cho, E.G, Park, Y.J. 2007. Power Core: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics.* 23: 2155-2162.

Knu¨pffer H and van Hintum Th J.L .1995. The barley core collection: an international effort. In: Hodgkin T, Brown AHD, van Hintum ThJL and Morales BAV (eds) Core Collections of Plant Genetic Resources. New York: International Plant Genetic Resources Institute (IPGRI), John Wiley & Sons, pp. 171–178.

Kramer, P.J. and Boyer, J.S. 1995. *Water Relations of Plants and Soils.* San Diego Academic Press. Pp. 697-701.

Kumar, J. and Abbo, S. 2001. Genetics of flowering time in chickpea and its bearing on productivity in Semi-arid environments. *Adv. Agron.* 72:107-138.

Konieczny, A. and Ausubel, F. M. 1993. A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* 4: 403-410.

Kupicha, F.K. 1981. Vicieae. Advances in legume, *systematics,* 1: 377–381.

Lagercrantz, U., Ellegren, H. and Andersson, L. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res* 21: 1111-1115.

Levitt, J. 1972. Responses of plants to Environmental Stresses. New York and London: Academic.

Lev-Yadun, S.,Gopher, A. and Abbo, S. 2000. The Cardle of Agriculture.Science.288:1602-1603

Liu, K. and Musa, S.V. 2005. Power marker: An integrated analysis environment for genetic marker analysis. *Bioinformatics*. 21: 9.

Lörz, H. and Wenzel, G**.** 2005. Molecular marker systems in plant breeding and crop improvement. Biotechnology in Agriculture and forestry, Springer, Berlin.

Marshall, D.R. 1989. Limitations to the use of germplasm collections. Pp. 105-120 *In:* The Use of Plant Genetic Resources (A.H.D. Brown, O.H. Frankel, D.R. Marshall and J.T. Williams, eds.). Cambridge University Press, Cambridge, UK.

Maxted, N., Forde-Lloyd, B.V. and Hawkes, J.G. 1997a. Complementary conservation strategies. In: Maxted, N., Forde-Lloyd, B.V. and Hawkes, J.G. (eds) *Plant Genetic Conservation: The* In Situ *Approach*. Chapman and Hall, London, pp. 15–39.

Mitra, J. 2001. Genetics and genetic improvement of drought resistance in crop plants. *Current Science*, 80: 758-763.

Mooney, H. A., Ehleringer, J. and Björkman, O. 1977. The energy balance of leaves of the evergreen desert shrub Atriplex hymenelytra. Oecologia, 29:301-310.

Muehlbauer, F. J. and Singh, K. B. 1987**.** Genetics of chickpea. Pages 99-125 *in* M. C. Saxena and K. B. Singh, eds. The chickpea. CAB International, London, UK.

Muehlbauer, F.J. and Tullu, A. 1997. *Cicer arietinum* L. *Purdue University:* New crop fact sheet. Retrievedfrom *http://www.hort. purdue.edu/newcrop/cropfactsheets/Chickpea. html*.

Mueller, U. G. and Wolfenbarger, L. L. 1999. AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution*. 14: 389-394.

Nordborg, M. and Tavare, S. 2002. Linkage disequilibrium: What history has to tell us? *Trends Genet.* 18*:* 83–90.

Nyquist, W. E. 1991. Estimation of heritability and prediction of selection response in plant populations. Critical Rev. *Plant Sci.* 10: 235-322.

Odong T.L., van Heewaarden J. 2011. Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data. *Theo.Appl.Gene* 123:195-205.

Orru`, L., Catillo, G., Napolitano, F., De Matteis, G., Scata` , M.C., Signorelli, F. and Moioli, B. 2009. Characterization of a SNPs panel for meat traceability in six cattle breeds. Food Control, 20: 856–860.

Pande, S., Siddique K.H.M., Kishore, G.K., Baaya, B., Gaur P.M., Gowda C.L.L., Bretag T., Crouch, J.H. 2005. *Ascochyta* blight of chickpea (*Cicer arientinum* L.): A review of biology, pathogenicity and disease management. Australian *J. Agri. Res*. 56:317-332.

Passioura, J. B. 1996. Drought and drought tolerance. Plant Growth Regulation 20: 79-83.

Paterson, A. H., Bowers, J. E., Burow, M. D., Draye, X., Elsik, C. G., Jiang, C. X., Katsar, C. S., Lan, T. H., Lin, Y. R. and Ming, R. 2000. Comparative genomics of plant chromosomes. *Plant cell* 12**:** 1523-1539.

Payne, K.W. Murray, D.A. Harding, S.A., Baird, D.B. and Soutar, D.M. 2009. Genstat for windows (12^{th} ed.) introduction. VSN International, Hemel Hempsteed.

Peeters, J.P. and J.A. Martinelli. 1989. Hierarchical cluster Analysis as a tool to manage variation in germplasm collections. *Theor. Appl. Genet*. 78:42–48.

Perez, T., Albornoz, J. and Dominguez, A. 1998. An evaluation of RAPD fragments reproducibility and nature. *Mol. ecol*. 7: 1347-1357.

Permesan and Yohe. 2003. A global coherent fingerprinting of climate change impacts across natural system. *Nature*. 421:2. 120-145.

Perrier, X. and Jacquemoud-Collet, J.P.2006.DARwin software http://darwin.Ciradfr/darwin

Polanco, C., Gonzalez, A. I., Puente, R. d. l., Somalo, S. and Ruiz, M. L. 2005. Cloning of AFLP markers detected by fluorescence capillary electrophoresis. *Plant molecular biology* 23: 271-277.

Pritchard, J.K., Stephens M, Rosenberg NA, Donnelly P. 2000b. Association mapping in structured populations. *Am J Hum* Genet 67:170-181.

Pritchard, J.K., M. Stephens and P. Donelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

Pritchard J.K., Wen W. 2007. Documentation for structure Software, version 2.2. Department of Human Genetics, University of Chicago, Chicago.

Prosser, J. 1993. Detecting single-base mutations. *Trends Biotechnol* 11: 238-246.

Quisenberry, J.E. 1982. Breeding for drought resistance and plant water efficiency. In *Breeding Plants for less favourable environments*. (Eds. M.N.Christiansen, C.P. Lewis) John Wiley Publishers: New York.

Rehman, A. U. 2009. Characterization and Molecular Mapping of Drought Tolerance in Kabuli Chickpea. PhD Thesis. Pp 5-22. Saskaton, Canada.

Rao, L. S., Usha Rani, P. Deshmukh, P. S., Kumar, P. A. and Panguluri, S. K. 2007. RAPD and ISSR fingerprinting in cultivated chickpea (*Cicer arietinum* L.) and its wild progenitor *Cicer reticulatum* Ladizinsky. Genet. Resour. *Crop Evol*. 54: 1235–1244.

Redden, B., Berger, J. 2007. History and origin of chickpea. In: Yadav SS, Redden B, Chen W, Sharma B (eds.). Chickpea Breeding and Management. CAB International, Wallingford. pp. 1-13.

Redden DT, Divers J, Vaughan LK, Tiwari HK, Beasley TM, Ferna´ndez JR, Kimberly RP, Feng R, Padilla MA, Liu N, Miller MB, Alliso D.B. 2006. Regional Admixture Mapping and Structured Association Testing: Conceptual Unification and an Extensible General Linear Model PLoS *Genetics*/ www.plosgenetics.org Vol. 2, Issue 8 | e137: 1254-1264.

Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., Kresovich, S., Goodman, M. M., and Buckler, E. S., IV. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *PNAS* **98**, 11479-11484.

Risch, N. and Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science* 273*:*1516–1517.

Rosielle, A.A. and Hamblin, J. 1981. Theoretical aspect of selection for yield in stress and non-stress environments. *Crop sci*., 21:943-946.

Sabaghpour, S. H., Mahmodi A.A., Saeed A., Kamel M., Malhotra R.S. 2006. Study on chickpea drought tolerance lines under dryland condition of Iran. *Indian J.Crop Sci*. 1(1-2): 70-73.

SAS Institute Inc**.** 2004. SAS Enterprise Guide 3.0.0.369.Cary, NC, USA

Schwarz, G., Herz, M., Huang, X. Q., Michalek, W., Jahoor, A., Wenzel, G. and Mohler, V. 2000. Application of fluorescence-based semi-automated AFLP analysis in barley and wheat. *Theor. appl. gene*. 100: 545-551.

Serraj, R. and Sinclair, J. R. 2002. Osmolyte accumulation: Can it really help increase crop yield under drought conditions? *Plant Cell Environ*. 25: 333-34.

Setter, J.L., Yan, J., Warburton, M., Ribaut, J-M., Xu, Y., Sawkins, M., Buckler, E.S., Zhang, Z. and Gore, M.A., 2011. Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. *J. exp. bota.* 62(2): 701-716.

Shriner D., Vaughan LK, Padilla MA and Tiwari H.K (2007) problems with genome wide association studies. *Science,* 316: 1840-1842.

Singh, K. B. 1997. Chickpea (*Cicer arietinum* L.). *Field Crops Research* 53:161-170.

Singh, K. B. 1987. Chickpea breeding. Pp. 127–162. *In* M.C. Saxena and K.B. Singh (ed.). The Chickpea. CAB International, Wallingford, UK.

Singh, K.B. and Ocampo B. 1997. Exploitation of wild *Cicer* species for yield improvement in chickpea. *Theor. Appl. Gene*. 95: 418-423.

Smale M. and McBride, T. 1996. Understanding global trends in the use of wheat diversity and international flows of wheat genetic resources. In CIMMYT World Wheat Facts and Trends - Part1 Pp 35-53.

Soltani, A., Khooie, F.R., Ghassemi-Golezani, K., Moghaddam, M., 2001. A simulation study of chickpea crop response to limited irrigation in a semiarid environment. *Agriculture water management*, 49:225-237.

Soltani, A., Hammer, G.L., Torabi, B., Robertson, M.J. and Zeinali, E. 2006. Modeling chickpea growth and development: Phonological development. *Field Crops Research*, 99, 1-13.

Soltis, D.E. and Soltis, P.S.1989.Polyploidy, breeding systems and genetic differentiation in homosporous pteridophytes. In: D.E. Soltis and Soltis, P.S.(eds), Isozymes in plant biology. Dioscorides Press, Portland, Oregon, 1-3.

Southern, E.M. 1975. Detection of specific sequence among DNA fragments separated by gel electrophoresis. *J. Mol. Bio*. 98: 503-517.

Suzuki, Y., Sekiya, T. and Hayashi, K. 1991. Allele-specific polymerase chain reaction: a method for amplification and sequence determination of a single component among a mixture of sequence variants. *Anal Biochem* 192: 82-84.

Syvanen, A. C. 2005. Toward genome-wide SNP genotyping. *Nat Genet* 37: 5-10.

Talebi, R., Fayaza, F., Maradi, M., Pirseyedi, S. M. and Naji, A. M. 2008. Genetic Relationships among chickpea elite lines based on RAPD and agronomic markers. *Int. J. Agri. Bio.*, 10: 3. Pp 301-315.

Tanksley, S. D., Young, N. D., Paterson, A. H. and Bonierbale, M. W. 1989. RFLP mapping in plant breeding: new tools for an old science. *Biotechnol* 7: 257-264.

Tanksley, S. D., Ganal, M. W. and Martin, G. B. 1995. Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends Genet*, 11: 63–68.

TASSEL 2.0.1, last update: January 30, 2007 North Carolina State University and Cornell University. WWW.maizegenetics.net/tassel.

Tautz, D. and Renz, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* 12**:** 4127-4138.

Thiel, T., Michalek, W., Varshney, R. K. and Graner, A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Gene.* 106: 411-422.

Thompson, J.A., R.L. Nelson, and L.O. Vodkin. 1998. Identification of diverse soybean germplasm using RAPD markers. *Crop Sci*. 38:1348–1355.

Tingey, S. V., Rafalski, J. A. and Hanafey, M. K. 1994. Genetic analysis with RAPD markers. NATO ASI ser., *Ser. H. Cell biol*. 81: 491-500.

Tohm, J., Jones, P., Beebe, S. and Ibanga, M. 1995. The combined use of agro ecological and characterization data to establish the CIAT *Phaseolus vulgaris* core collection. Pp. 95–108. In T. Hodgkin *et al*. (ed.). Core collections of plant genetic resources. *Int. Plant Genet. Resour*. Inst. John Wiley & Sons, New York.

Thornsberry J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E. S. 2001. Dwarf polymorphisms associate with variation in flowering time. *Nat. Genet.* 28*:* 286–289.

Tommasini, L., Schnurbusch, T., Fossati, D., Mascher, F., and Keller, B. 2007. Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties. *Theor Appl Genet* **115**, 697-708.

Turner, N. C. and Begg, J. E. 1978. Characterization of drought tolerance in chickpea: In Plant Relations in Pastures (J.R.Wilson, Ed), Pp. 50-66. Melbourne: CSIRO.

Turner N.C. 1979. Drought resistance and adaptation to water deficits in crop plants. Pages 343-372 in H. Mussell and R.C. Staples. Eds. Stress Physiology of crop plants. Wiley. New York.

Upadhyaya, H. D., Dwivedi, S. L., Baum, M., Varsheny, R. K., Udupa, S. M., Gowda, C.L.L., Hoisington, D. and Singh, S. 2001. Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant. Biol.* 8:106.

Upadhaya, H. D., Dwivedi, S. L., Gowda, C.L.L. and Singh, S. 2007. Identification of diverse germplasm lines for agronomic traits in chickpea (*Cicer arietinum* L.) core collection for use in crop improvement. *Field Crops Res*., 100: 320–6

Upadhyaya, H. D and Ortiz, R. 2001. A minicore subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor and Appl Gene.*, 102: 1292-1298.

Upadhyaya HD, Gowda CLL and Sastry DVSSR. 2008. Plant genetic resources management: collection, characterization, conservation and utilization. Journal of SAT Agricultural Research 6.

van der Maesen, L. J. G. 1972. Cicer L. Monograph of the Genus with Special Reference to the Chickpea (Cicer *arietinum* L.) Veeman and Zoren, Wageningen, 72: 342.

van der Maesen, L. J. G. 1987. Origin, history, and taxonomy of chickpea. P.11-34. In: M.C. Saxena, and K. B. Singh, (eds), The Chickpea. CAB International Publications, UK.

van der Maesen L.J.G, Maxted N., Javadi F, Coles S& Davies AMR (2007). Taxonomy of the genus Cicer revisited In: Chickpea breeding and management. Ed. S.S. Yadav, Indian Agricultural Research Institute, New Delhi, India. R. Redden, Victoria Institute for dry land Agriculture, Horsham, Australia. W Chen, USDA-ARS, Washington State University, USA and B Sharma, Indian Agriculture Research Institute, New Delhi, India.

Varshney, R. K., Hiremath, P. J., Lekha, P., Kashiwagi, J., Balaji, J., Deokar, A.A., Vadez, V., Xiao,Y., Srinivasan, R., Gaur, P.M., Siddique, K. H. M., Town, C.D. and Hoisington, D. A. 2009. A comprehensive resource of drought and salinity responsive ESTs for gene discovery and marker development in chickpea. *BMC Genomics*, 10:523.

Vavilov, N.I. 1926. Studies on the origin of cultivated plants. Leningrad. pp. 129-238

Verslues, P. E., Agarwal, M., Katiyar-Agarwal, S. and Zhu, J. 2006. Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *The Plant J.* 45: 523-539.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23: 4407-4414.

Wang Y., Rannala B. 2005. *In silico* analysis of disease-association mapping strategies using the coalescent process and incorporating ascertainment and selection. *Am J Hum Genet* 76:1066–1073

Wang, J.C., Hu, J., Xu, H.M., Zhang, S. 2007. A strategy on constructing core collections by least distance stepwise sampling, *Theor. Appl. Gene.* 115: 1-8

Waugh, R. and Powell, W. 1992. Using RAPD markers for crop improvement. *Trends Biotechnol.* 10: 186-191.

Weir, B. S. 1990. Genetic data analysis. *American Journal of Human genetics*, 52: 506-515.

Welsh, J. and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18: 7213-7218.

Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A. and Kilian, A. 2004. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci.* USA. 101: 9915-9920.

Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18: 6531-6535.

Wilson, L. M., Whitt, S. R., Ibáñez, A. M., Rocheford, T. R., Goodman, M. M., and Buckler, E. S. IV. 2004. Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Analysis. *The Plant Cell* 16: 2719-2733.

Winter, P., and Kahl, G. 1995. Molecular marker technologies for crop improvement. *World J. Microbiol. Biotechnol*. 11:449-460.

Xiong M., Jin L. 1990. Comparison of the power and accuracy of biallelic and microsatellite markers in population based gene mapping methods. *Am. J. Hum. Genet*. 64: 629-640

Xu, X., Peng, M., Fang, Z. and Xu, X. P. 2000. The direction of microsatellite mutations is dependent upon allele length. *Nature Genet*. 24: 396–9.

Xu, J.L., Lafitte, H.R., Gao, Y.M., Fu, B.Y., Torres, R., and Li Z.K. 2005. QTLs for drought escape and tolerance identified in a set of random introgression lines of rice. *Theor. Appl. Genet*. 111:642–1650.

Xu, H., Mei, H., Hu, J., Zhu, J. and Gong, P. 2006. Sampling a core collection of Island cotton (*Gossypium barbadense* L.) based on the genotypic values of fiber traits. *Gene. Resour. and Crop Evo.* 53: 515-521.

Yu, J. and Buckler, E.S. 2006. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol*. 17*:* 155–160.

Zarea-Fizabady A, Ghodsi M. 2004. Evaluation of yield and yield components of facultative and winter bread wheat genotypes (*Triticum aestivum L.)* under different irrigation regimes in Khorasam province in Iran. *J. Agron.*, 3 (3): 184-187.

Zhu C., Gore M.A., Bucklers E.S., Yu J. 2008. Status and prospect of association mapping in plants. *Plant Genome* 1: 5-20.

8. Appendix

Appendix 1. Sequence information of SNP markers

Sr no.	**Markers name**	**Allele Y (1)**	**Allele X (2)**	**Complete sequence**
1	CKaM0003	G	T	GGAATAAGTGTCTCT[T/G]CCCTGGATAATTTGT
2	CKaM0017	A	G	TTAGGCCTTCCAGTG[G/A]TATCCCATTTATCAC
3	CKaM0025	C	T	GTACAAATGTTTTTC[T/C]TAAATTTTACTACTT
4	CKaM0033	T	**G**	TTTTTGCGTTTGCTT[G/T]TATCTTGCAATGGAA
5	CKaM0042	A	G	GGAACCACCGGTGGT[G/A]GCTATGGTGGAGGAT
6	CKaM0043	T	C	GGTGGTGNATGGATA[C/T]GGTGACAAATACTCC
7	CKaM0167	C	T	ATGACCAATGTTAAT[T/C]TGAACTGAAGCGTGA
8	CKaM0204	C	T	AAGGAGACATCAACA[T/C]CGACGCCAATTTCAA
9	CKaM0234	C	T	ATCATGAAGTAAAGA[T/C]GCAACACAAATAAGA
10	CKaM0249	C	T	TGTTGATAATAATCT[T/C]GCAGTGCTCTCCTGA
11	CKaM0290	G	C	TCTCTTTACTCGGCC[C/G]ACCTAGATCACACTC
12	CKaM0317	C	T	GCCTTAAAACCCTTA[T/C]ATTGCAAAGCACAAT
13	CKaM0321	C	T	AATATTCCGGATTGT[T/C]CCATTTGGGCTCTTC
14	CKaM0343	C	T	ACAGGTTACTAAATG[T/C]ATCTGATTTGCAGTG
15	CKaM0405	G	A	TTCTTGAAAGGGAAA[A/G]GGTCTCAAGGGTTTG
16	CKaM0411	C	A	AGCCATACTAATGCC[A/C]GCATGGATGAGATTC
17	CKaM0447	G	A	TCATGGTCCAGTTAG[A/G]ACAAATGGTGGATGG
18	CKaM0462	G	A	GCACCAAATACTTGC[A/G]CTAAACTTTGACGGT
19	CKaM0477	C	T	ATAATATCAGTTGTA[T/C]GTGCTATGTATTGAG
20	CKaM0493	C	T	GTTATAGAATGCAGC[T/C]TGTGCCTCTACAGGA
21	CKaM0526	G	C	AATCATCAAATTTTT[C/G]AAGTTTGTCCATGAA
22	CKaM0588	T	A	GTGCATTTTATGGTA[A/T]TCATGTGCTAGGGAG
23	CKaM0612	C	T	TCAGAAGAAAACTGC[T/C]TTGAATCGGCTGGTT
24	CKaM0630	G	A	TGGACAAAGTGAAGA[A/G]CAGGCTACACTGAAC
25	CKaM0639	G	T	TTTTTTCACAGCTGT[T/G]AGTGCCACCAACCTT
26	CKaM0647	G	A	TGTTGAGTTGCTTTT[A/G]TTAGTTTTTCCAAGT
27	CKaM0657	G	T	TTCTTACACTCTATT[T/G]GTTCATTGTGTGTAG
28	CKaM0707	G	A	TCACATTTCATCCCA[A/G]TTGTGAAACAAGTTT
29	CKaM0722	G	C	TATGTATAGGAGTTT[C/G]TGTCTGTATGTAATT

30	CKaM0723	A	G	AATGTTTCAAAATAT[G/A]TTACAACAATTTCAC
31	CKaM0750	T	C	ATCAAATTACACACC[C/T]GACTTTCATCTCAAA
32	CKaM0804	T	C	AACACTTGGAGATGC[C/T]CTTATCAGCCGCCTC
33	CKaM0888	C	T	CTATCTAGGTATCAG[T/C]TTTCTTTCACATCAC
34	CKaM0993	C	T	CTAGACACTGGATTC[T/C]GCATTGTGTAGTGGA
35	CKaM0999	A	T	TGGAGCATATGTTAT[T/A]GTGACGGATTATGGT
36	CKaM1101	C	T	CCCTCCCTTCCCTTC[T/C]CATCCCTCTCTACCA
37	CKaM1140	G	T	AATTACATTCTTCAA[T/G]GTGAAAAATTGACCT
38	CKaM1175	G	A	GCGGTGCGAACTATG[A/G]AACCGTCGCTATAGA
39	CKaM1190	T	C	GTGAAATTGTTGTAA[C/T]ATATTTTGAAACATT
40	CKaM1254	A	G	AATGTTTCAAAATAT[G/A]TTACAACAATTTCAC
41	CKaM1293	C	T	GGCAGAAATTGAATG[T/C]GATTCTTTATTGCTA
42	CKaM1317	C	A	GTAAATATAGGCTAT[A/C]CTTCAACTCAATGTG
43	CKaM1328	G	A	CTTGATTGGTGCTAA[A/G]TGTTAAAGTCCAGCA
44	CKaM1356	A	G	TTTGAGATGAAAGTC[G/A]GGTGTGTAATTTGAT
45	CKaM1641	G	A	GTAGTGGTGTCTTCC[A/G]TTAGCTATTTTGGTA
46	CKaM1651	A	G	TTTATTGAAGATAAG[G/A]TTGCTGATAATGCTG
47	CKaM1788	T	C	ATCAAATTACACACC[C/T]GACTTTCATCTCAAA
48	CKaM1832	T	C	ATCAAATTACACACC[C/T]GACTTTCATCTCAAA
49	CKaM1842	C	T	TGCTGGACTTTAACA[T/C]TTAGCACCAATCAAG
50	CKaM1848	C	G	ACTGGAATTCCTAAT[G/C]AATTTGTAAGACTCT
51	CKaM1902	A	G	TGGAATATATCAAGT[G/A]CCGCAACTCATTAAC
52	CKaM1903	T	C	CAGCATTATCAGCAA[C/T]CTTATCTTCAATAAA
53	CKaM1933	G	A	AAGTTGAACTGCTAT[A/G]CTCAAAAGGTTGACA
54	CKaM1963	C	T	GACTCCTGTATTATT[T/C]AGCGTCATACGGTCA
55	CKaM1971	G	A	TTGTTAGTGGTGATG[A/G]GAAAAGGGTTAAGCG